70427·

Access DB# _____ (192)

SEARCH REQUEST FORM
Scientific and Technical Information Center

Requester's Full Name: RITA MITRA   Examiner #: 77995 Date: 7/8/02
Art Unit: 1653   Phone Number 30 605-1211   Serial Number: 09/691344
Mail Box and Bldg/Room Location: 9B01/CM1 Results Format Preferred (circle): PAPER DISK E-MAIL
Rm 9B.03

If more than one search is submitted, please prioritize searches in order of need.
****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: "Novel human proteins and polynucleotides encoding the same".

Inventors (please provide full names): Gregory Donoho, C. Alexander Turner, Michael Nehls, Glenn Fric-drich, Brian Zambrowicz, Arthur Sands.
Earliest Priority Filing Date: 10/19/1999

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

I would request an expedited literature (both patent and Non-patent) search on above application because this is a date case. Don't do Sequence the search.

The search should encompass nucleotides that encode proteins, wherein proteins have structural similarity to animal CUB domain proteins, coagulation factors V and XIII, milk fat globule-EGF factor 8, transcriptional repressor AE-binding protein-1 and neurophilins 1 and 2.

Keywords:
CUB domain, discoidin domain, NHP, agonist, antagonist

1604-56
15.46 repressor.
6 tamoxifen transcriptional activity human

Mary Hale – Supervisor, Info. Branch
STIC - Biotech/Chem. Library
CM-1 Room E01
703-308-4258

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: Mann	NA Sequence (#)_____	STN 120.70
Searcher Phone #: _____	AA Sequence (#)_____	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: _____	Bibliographic ✓	Dr.Link _____
Date Completed: 7/9	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: 18	Other _____	Other (specify)_____

PTO-1590 (1-2000)

```
Db 1621 gaggctccagaaaacagaactgttggaagcagagcagagattaatatcccttgccagaatcagtca
Db 1681 gctcaggttcaccatccagctaatgataatgagagaagagagatgacacaaaagttagatccatc 1740
Qy 1681 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 1740
Db 1681 gctcaggttcaccatccagctaatgataatgagagaagagagatgacacaaaagttagatccatc 1740
Qy 1741 acaagtgtatggcaggttaa 1761
        ||||||||||||||||||||
Db 1741 acaagtgtatggcaggttaa 1761
```

RESULT 2

AAS00660

```
ID   AAS00660 standard; cDNA, 3594 BP.
XX
AC   AAS00660;
XX
DT   07-SEP-2001 (first entry)
XX
DE   Human TANGO 229 cDNA.
XX
KW   Human; TANGO 229; T cell; heart; liver; pancreas; brain; lung;
KW   skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW   bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW   anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW   malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW   attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW   hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW   muscular dystrophy; immuno-competence; vertebrate; blood; serum.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            72..2219
FT                  /tag= a
FT                  /product= "Human TANGO 229"
FT   sig_peptide    72..173
FT                  /tag= b
FT   mat_peptide    174..2216
FT                  /tag= c
FT                  /product= "Mature human TANGO 229"
XX
XX
PN   WO200129088-A1.
XX
PD   26-APR-2001.
XX
PF   23-JUN-2000; 2000WO-US17386.
XX
PR   19-OCT-1999; 99US-0420707.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   MacKay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
XX
DR   WPI; 2001-308477/32.
DR   N-PSDB; AA000670.
XX
```

New isolated nucleic acid molecule for diagnosis, prevention, and therapy of human and other animal disorder, or as modulating agent for regulating cellular processes.

Claim 1; Fig 1; 263pp; English.

The sequence represents a cDNA which encodes human TANGO 229 polypeptide. This protein and similar others exhibit the ability to affect growth, proliferation, survival, differentiation, activity, morphology, or movement/migration of, e.g. T cells and cells of the heart, liver, pancreas, brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral blood leukocyte, bone marrow or thymus tissue. They can be used as modulating agents for regulating cellular processes, thus, the proteins and their associated nucleic acids can be used to prognosticate, prevent, diagnose, or treat disorders associated with physiological processes. These disorders include abnormal blood coagulation, asthma,

```
CC   anaphylaxis, hepatitis, multiple sclerosis, cancer, coronary artery
CC   disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis,
CC   meningitis, attention deficit disorder, Crohn's disease, gastroenteritis,
CC   goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary
CC   embolism and muscular dystrophy. Antibodies to disorders such as these
CC   can be made by providing a polypeptide of the invention to an
CC   immuno-competent vertebrate and harvesting blood or serum from the
CC   vertebrate.
XX
SQ   Sequence 3594 BP; 958 A; 810 C; 921 G; 905 T; 0 other;
```

```
Query Match      95.8%;  Score 1686.8;  DB 22;  Length 3594;
Best Local Similarity   99.9%;  Pred. NO: 0;
Matches 1688;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy   71 ggcccgggcagctcgcgcgctcggggggtcccggcgcgggggagggggagggccgagcttgccaagctggcgc 130
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ggcccgggcagctcgcgcgctcggggggtcccggcgcgggggagggggagggccgagcttgccaagctggcgc 60

Qy  131 ccagcggggtcatgtcgagtctcgggtcccggcgcccggccgcggggcggggccactcggcgcggcgccggc 190
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ccagcggggtcatgtcgagtctcgggtcccggcgcccggccgcggggcggggccactcggcgcggcgccggc 120

Qy  191 ggggcccctggcttgtctcgcggtctccgcccgccccgcgcggccgtcgcaggcgggaggggc 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ggggcccctggcttgtctcgcggtctccgcccgccccgcgcggccgtcgcaggcgggaggggc 180

Qy  251 tgggtcgatgggctgtcggacaaccttacttacttcaaggtagtggcacatggcaagatactaaga 310
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 tgggtcgatgggctgtcggacaaccttacttacttcaaggtagtggcacatggcaagatactaaga 240

Qy  311 atattccccgggaccaccccaatcacacgtgttccgaaaagcaaattacgtaccaaagg 370
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 atattccccgggaccaccccaatcacacgtgttccgaaaagcaaattacgtaccaaagg 300

Qy  371 ggaaaagactgaattcctgagagttggggagattcgttgatatcgaatcccagaccctgctctcg 430
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ggaaaagactgaattcctgagagttggggagattcgttgatatcgaatcccagaccctgctctcg 360

Qy  431 actacatcctctctcaccagatccttcaggatggtccatacctgtggaagtatgactg 490
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 actacatcctctctcaccagatccttcaggatggtccatacctgtggaagtatgactg 420

Qy  491 tcccaagagaacttcttgttgaacacaagtaaccgtccgcttgagagtggatcc 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 tcccaagagaacttcttgttgaacacaagtaaccgtccgcttgagagtggatcc 480

Qy  551 acatttcctggccggggttttttggctgaccatgccgagcacaccgagattaataa 610
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 acatttcctggccggggttttttggctgaccatgccgagcacaccgagattaataa 540

Qy  611 catgttggaacgagctagccccatttattttgtgtgacctatcgcagcgccatccagattttaataa 670
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 catgttggaacgagctagccccatttattttgtgtgacctatcgcagcgccatccagattttaataa 600

Qy  671 gttgtagacgtagcagagagacaactttctgggaatatggtagatgatatagaatacct 730
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 gttgtagacgtagcagagagacaactttctgggaatatggtagatgatatagaatacct 660

Qy  731 cttatattgtgcaaagtgccatcctccatgaggaataaactgctgatgaactagttggccaga 790
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 cttatattgtgcaaagtgccatcctccatgaggaataaactgctgatgaactagttggccaga 720

Qy  791 tcagtgtgcttcagcgacaaaagggcagtcgatatgaaggatggtctgaggattggttgttc 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 tcagtgtgcttcagcgacaaaagggcagtcgatatgaaggatggtctgaggattggttgttc 780

Qy  851 tttcgaaggatggtccctgcaagcacaagcgattctgttacctcaatggttgcagca 910
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 tttcgaaggatggtccctgcaagcacaagcgattctgttacctcaatggttgcagca 840

Qy  911 gatcctgagctttgaacctgacgggcaaatcagaagcttcttcctctaatgtcaagtcggtca 970
```

RESULT 3
AAS00615

```
ID   AAS00615 standard; cDNA; 1620 BP.
XX
AC   AAS00615;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Novel human protein (NHP) DNA sequence #3.
XX
DE   Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
     mostly; high blood pressure; connective tissue disorder; infertility;
     mediated pathway; ss;
```

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1620
FT                   /tag= a
FT                   /product= "Novel human protein #3"
XX
PN   WO00129219-A1.
XX
PD   26-APR-2001.
XX
PF   08-OCT-2000; 2000WO-US28798.
XX
PR   19-OCT-1999; 99US-0160285.
PR   18-FEB-2000; 2000US-0183583.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
DR   WPI; 2001-290917/30.
DR   P-PSDB; AAU00630.
XX
PT   Novel nucleic acid encoding human CUB-domain containing protein, useful
PT   for drug screening, diagnosis and treatment of physiological disorders
PT   or diseases -
XX
PS   Disclosure; Page 28-29; 33pp; English.
XX
CC   The sequence represents a polynucleotide which encodes a novel human
CC   protein (NHP) containing a CUB domain (an extracellular domain). CUB
CC   proteins have been associated with regulating development, modulating
CC   cellular processes and preventing infectious disease. NHP nucleotide
CC   sequences are useful for gene therapy of physiological disorders or
CC   diseases. NHP oligonucleotides are useful as hybridisation probes for
CC   screening libraries and assessing gene expression. NHP nucleotide
CC   are useful for detecting mutant or inappropriately expressed NHPs (for
CC   example, those proteins associated with obesity, high blood pressure,
CC   connective tissue disorders and infertility) for the diagnosis of a
CC   disease. The polynucleotides may also be used in screening for drugs
CC   effective in the treatment of symptomatic or phenotypic manifestations of
CC   perturbing the normal function of NHP in the body. Nucleotide constructs
CC   encoding NHP products are used to genetically engineer host cells to
CC   express such products in vivo. These host cells allow for the
CC   identification of compounds that bind to NHP receptors or trigger
CC   NHP-mediated pathways.
XX
SQ   Sequence 1620 BP; 456 A; 353 C; 422 G; 389 T; 0 other;
```

```
Query Match       92.0%;  Score 1620;  DB 22;  Length 1620;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1620;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy     142 atggtgccggccgccgccgccgggccggcgcactggccgccgggccgggcccctg 201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atggtgccggccgccgccgccgggccggcgcactggccgccgggccgggcccctg  60

Qy     202 gcttgctgctgcagctccggccgccgggctgcaggcgcggaagacgctgggtgatcggc 261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gcttgctgctgcagctccggccgccgggctgcaggcgcggaagacgctgggtgatcggc 120

Qy     262 tgtggacacctagtgacttatcaggatagtgcacaatgaactcctaagaattatcccggc 321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tgtggacacctagtgacttatcaggatagtgcacaatgaactcctaagaattatcccggc 180

Qy     322 accttacccaatcacactgtttgccaaagacaatacacagtaccaaaggagaaaagacta 381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 accttacccaatcacactgtttgccaaagacaatacacagtaccaaaggagaaaagacta 240

Qy     382 atttctgaggttgggagattcgaatcccagaaccagtgtgcttctgactactctcc 441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          atttctgaggttgggagattcgaatcccagaaccagtgtgcttctgactactctcc 441
```

```
Db     841 gatccttgagttcttgaacctgacgggcaaatcagagcttcttcctcctcgtcgtca 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy         ...
Db     901 atgagagtcggagaccaagttcaactggctcactgtggccaagcaacctcaggagccaaggcc 960
Db     961 catcatggggcttcggggcgacagtgacaactaacacaagccaacaaggtggctgagatcg 1020
Db    1021 atttgggaggagaaaagaaaagaaataacacaggaattaggcacacgagatcataacagtc 1080
Db    1081 tcaacttttatgttaagagttctgtgatgaacttcaaaaacaataattctaagtcaagact 1140
Db    1141 cctataaaggaattctgaatatgaagaaaggtgttcaagagttatggctgaactc 1200
Db    1201 accagtgcaaaacaattcatcccctcctcgtggccagatatgtcgggttgtccccc 1260
Db    1261 agacacagcaccaaggtagcctgaggtggagctcattggttgccagattacacaag 1320
Db    1321 gtaatgtgatcatcggtsgtgcgcaggacaagtcaaagcaccagtgttccaactaagaaag 1380
Db    1381 aagtgagagacaatcacaaggccacaccccccggaagaaacatccacaggagaataaacatta 1440
Db    1441 caacggtgtgctattcctggtgctccctgttcctgggtgtttgctgctggatgggatct 1500
Db    1501 ttgcagccttagaaaaagaagaaagagagaagtccgtatggatgacgaaggctcaga 1560
Db    1561 aaacagactgttggaagcagaattaaatatccctttgcagacacattcagtcagagtta 1620
```

```
Qy    1621 cctcagctatgataatgaaggaggagctggagcacacaaatattagattctcaccacagtgata 1680
Qy    1681 tggcagatta 1690
Qy    1751 tggcaggtta 1760
```

[handwritten:] SEQ ID NO: 4
Database: A-Geneseq-032802
AC NO: AAU00670

```
Qy   280  GSLSDKRFLFTSNGCGSRSLSFEPDGQIRASSWQSVNESGDGQVHWSPGQARLQDQGPSWA  347
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  gslsdkrflfftsngcgsrslsfepdgqirasswqsvnesgdqvhwspgqarlqdqgppswa  300

Qy   348  SGDSSNNHKPREWLEIDLGEKKKITGIRTTGSTQSNFNFYVKSFVWNFKKNNSKWKTYKG  407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  sgdssnnhkprewleidlgekkkitgirttgstqsnfyvksfvmnfkknnskwktykg  360

Qy   408  IVNNEEKVFQGNSNFRDPVQONNFIPIVARYVRVVPQTWHQRIALKVELIGCOITTQGNDS  467
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  ivnneekvfqgnsnfrdpvqnnfipivaryvrvvpqtwhqrialkveligcqittqgnds  420

Qy   468  LVWRKTSQSTSVSTKKEDETTTRPIPSEETSTGINITTVAIPLVLLVLVFAGMGIFAAF  527
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  lvwrktsqstsvstkkedetttrpipseetstginittvaiplvllvlvfagmgifaaf  480

Qy   528  RKKKKKGSPYGSAEAOKTDCWKQIKYPFARHQSAEFTISYDNEKEMTQKLDLIITSDMAG  586
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  rkkkkkgspygsaeaqktdcwkqlkypfarhqsaeftisydnekemtqkldliitsdmag  539
```

```
RESULT   3
ID   AAU00670 standard; Protein; 715 AA.
AC   AAU00670;
XX
DT   07-SEP-2001  (first entry)
XX
DE   Human TANGO 229 polypeptide.
XX
KW   Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
KW   skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW   bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW   anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease;
KW   malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW   attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW   hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW   muscular dystrophy; immuno-competence; vertebrate; blood; serum.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   Peptide         1..34
FT                   /note="Signal peptide"
FT   Domain          35..455
FT                   /note="Extracellular domain"
FT   Protein         35..715
FT                   /note="Mature human TANGO 229"
FT   Domain          456..480
FT                   /note="Transmembrane domain"
FT   Domain          481..715
FT                   /note="Cytoplasmic domain"
XX
DR   WO200129088-A1.
XX
PN   26-APR-2001.
XX
PD   23-JUN-2000; 2000WO-US173386.
XX
PF   19-OCT-1999; 99US-0420707.
PR
PA   (MILL-) MILLENNIUM PHARM INC.
PI   Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
XX
DR   WPI: 2001-308477/32.
DR   N-PSDB; AAS00660.
XX
PT   New isolated nucleic acid molecule for diagnosis, prevention, and
PT   therapy of human and other animal disorder, or as modulating agent for
PT   regulating cellular processes
```

[handwritten:] SEQ ID NO: 4
Database: A-Geneseq-032802
AC NO: AAU00670

```
XX
PS   Claim 8; Fig 1; 263pp; English.
XX
CC   The sequence represents human TANGO 229 polypeptide. This protein and
CC   similar others exhibit the ability to affect growth, proliferation,
CC   survival, differentiation, activity, morphology, or movement/migration
CC   of, e.g. T cells and cells of the heart, liver, pancreas, placenta,
CC   brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
CC   blood leukocyte, bone marrow or thymus tissue. They can be used as
CC   modulating agents for regulating cellular processes, thus, the proteins
CC   and their associated nucleic acids can be used to prognosticate, prevent,
CC   diagnose, or treat disorders associated with physiological processes.
CC   These disorders include abnormal blood coagulation, asthma, anaphylaxis,
CC   hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
CC   atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
CC   deficit disorder, Crohn's disease, gastroenteritis, goitre,
CC   hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
CC   muscular dystrophy. Antibodies to disorders such as these can be made by
CC   providing a polypeptide of the invention to an immuno-competent
CC   vertebrate and harvesting blood or serum from the vertebrate.
XX
SQ   Sequence 715 AA;
```

```
Query Match     92.1%;  Score 2812;  DB 22;  Length 715;
Best Local Similarity  100.0%;  Pred. No. 2.8e-243;
Matches  538;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Qy    48  MVPGARGGALARAARGRLLALLLAVSAPLRLQAERLGDCCGHLVTYQDSGTWMSKNYPG  107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  mvpgargggalaraagrglallllavsaplrlqaeelgdgcghlvtygdsgtmtskypg   60

Qy   108  TYPNHTVCEKITVPFKGKRHLIRLGDLDLIESQTCASDYLLFSSSDQYGYCCGSMTVPKE  167
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  typnhtvcekitvpfkykrllrlsgdldlesqtcasdylflfsssdqy9pycgsmtvpke  120

Qy   168  LLLNTSEVTVRFESGSHISGRGGFLLTYASSDDPDLICLERASHYLRTEYSKFCPAGCRD  227
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  lllntsevtvrfsesghisgrgfflltyassdhpdllicleraswhylktesyskfcpagcrd  180

Qy   228  VAGDISGNMVDGYRDYSLLCKRAAIHAGIIADELGGOISVLQRKGISRYEGILANGVLSRD  287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  vagdisgnmvdgyrdtsllckaalhagiiadelggqisvalqrkgisryeglilangvlsrd  240

Qy   288  GSLSDKRFLFTSNGCSRSLSFEPDGQIRASSWQSVNESGDGQVHWSPGQARLQDQGPSWA  347
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  gslsdkrflfftsngcsrslsfepdgqirasswqsvnesgdqvhwspgqarlqdqgppswa  300

Qy   348  SGDSSNNHKPREWLEIDLGEKKKITGIRTTGSTQSNFNFYVKSFVWNFKKNNSKWKTYKG  407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  sgdssnnhkprewleidlgekkkitgirttgstqsnfyvksfvmnfkknnskwktykg  360

Qy   408  IVNNEEKVFQGNSNFRDPVQONNFIPIVARYVRVVPQTWHQRIALKVELIGCOITTQGNDS  467
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  ivnneekvfqgnsnfrdpvqnnfipivaryvrvvpqtwhqrialkveligcqittqgnds  420

Qy   468  LVWRKTSQSTSVSTKKEDETTTRPIPSEETSTGINITTVAIPLVLLVLVFAGMGIFAAF  527
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  lvwrktsqstsvstkkedetttrpipseetstginittvaiplvllvlvfagmgifaaf  480

Qy   528  RKKKKKGSPYGSAEAOKTDCWKQIKYPFARHQSAEFTISYDNEKEMTQKLDLIITSDMA  585
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  rkkkkkgspygsaeaqktdcwkqlkypfarhqsaeftisydnekemtqkldliitsdma  538
```

```
RESULT   4
ID   AAU00628 standard; Protein; 487 AA.
AC   AAU00628;
XX
DT   29-AUG-2001  (first entry)
```

RESULT 9

ID   AAH3673G
XX
AC   AAH3673G standard; cDNA; 252 BP.
XX
DT   03-SEP-2001 (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:3818.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US265224.
PR   29-SEP-1999; 9US-0157137.
PR   03-NOV-1999; 9US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG77329.
XX
PT   Nucleic acids encoding 1277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 5674-5675; 5903pp; English.
XX
CC   AAH33195 and AAF73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention,
CC   diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence, listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 252 BP; 68 A; 56 C; 68 G; 47 T 13 other;

Query Match   7.2%; Score 127.2; DB 22; Length 252;
Best Local Similarity   81.1%; Pred. No. 6.9e-23;
Matches 189; Conservative   0; Mismatches 39; Indels   5; Gaps   4;

Qy 1206 gaagacctataaaggaattgtgaataa-tgaagaaaagacgttcaggtaactctact 1264
          |||||||||||||| |||||||||||| |||||||||| |||||||||||||||||||
Db   19 ggggacctataaaggaattgtgagtaagagaataaaagagttcaggggaactctaact 78

Qy 1265 ttcgggaccagtgcaaaacaaattcatccctccaatcgttcggccaagatgt-gcggggtt 1323
          |||||||||||| |||||| || |||||||||||||||||||||||||||||| |||| |||
Db 1265 cggagaccacacatgcgacngcagnncataacctccaatngtggcogagatatgtcgcggatt 138

RESULT 10

ID   AAZ32182
XX
AC   AAZ32182 standard; cDNA; 6893 BP.
XX
DT   13-JAN-2000 (first entry)
XX
DE   Human coagulation factor V nucleotide sequence.
XX
KW   Human; coding sequence polymorphism; vascular pathology gene;
KW   polymorphic site; phenotype correlation; forensic; paternity testing;
KW   medicine; genetic analysis; vascular disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO9950454-A2.
XX
PD   07-OCT-1999.
XX
PF   26-MAR-1999;
PR   01-APR-1998; 98US-0054272.
XX
PA   (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Lander ES, Daley GO, Cargill M, Ireland JS, Rozen SG;
XX
DR   WPI; 1999-620066/53.
DR   P-PSDB; AAY49563.
XX
PT   Determination of polymorphisms in genes, especially those identifying
PT   predisposition to vascular disease -
XX
PS   Claim 1; Fig 26; 134pp; English.
XX
CC   AAZ32159 to AAZ32194 represent reference alleles for specifically
CC   claimed nucleic acid sequences from the present invention which comprise
CC   polymorphic sites as given in a table in the specification, selected
CC   from 92 single nucleotide polymorphisms in which the nucleotide at the
CC   polymorphic site is different from that at the same site in a
CC   reference allele. The nucleic acids, and primers and probes, are used to
CC   identify polymorphisms which may predispose an individual to disease,
CC   especially a vascular disease. They can also be used in phenotype
CC   correlations, forensics, paternity testing, medicine or genetic
CC   analysis. AAY49550 to AAY49573 represent the proteins which correspond
CC   to some of the reference alleles.
XX
SQ   Sequence 6893 BP; 2090 A; 1700 C; 1423 G; 1680 T; 0 other;

Query Match   5.4%; Score 94.4; DB 20; Length 6893;
Best Local Similarity   51.2%; Pred. No. 8.7e-14;
Matches 249; Conservative   0; Mismatches 231; Indels   6; Gaps   1;

Qy   898 aatggttgcagcagatccttgagtttgaacctgacgggccaaaatcagagctctcctca 957
Db 6266 aatggatgttccacacccctgggttatggaaaatgggaaagdatgaaaacaagcaaatcaca 6325

Qy   958 tggcagtcggtcaatgagagtgggagaccaagttcactgttcctcggccaagcgccgact 1017
Db 6326 gctctcgtttaagaaatcttggtcgggggaggattactggagaaccttccgtgccgctctg 6385

Qy 1018 caggaccaaggcccatcatcgggtcggcggacgagtagcaacaaccacacaaacacgagag 1077

handwritten: SEQ ID NO: 3 Database; N-Gene249-032802 Ac No; AAZ32182

RESULT 11

ID AAT03920 standard; cDNA; 6909 BP.
XX
AC AAT03920;
XX
DT 20-DEC-1995 (first entry)
XX
DE Human Factor-V wild-type cDNA.
XX
DE Factor-V; thrombosis; thrombophilia; diagnosis; anticoagulant;
KW activated protein-C; APC; homozygosity; heterozygosity; ss.
XX
KW Bertina RM, Reitsma PH;
KW (UYLE') RIJKSUNIV LEIDEN.
XX
OS Homo sapiens.
XX
PN WO9521938-A1.
PD 17-AUG-1995.
XX
PR 14-FEB-1994; 94EP-020Q377.
PF 14-FEB-1995; 95WO-EP00553.
XX
DR WPI; 1995-293134/38.
XX
PT Screening for genetic defect associated with thrombosis and/or poor
PT anticoagulant response to activated protein C - useful to determine
PT homozygosity or heterozygosity for a mutation in Factor V, Va, VIII
PI or VIIIa.
XX
PS Claim 16; Page 60-65; 98pp; English.
XX
CC The probes/primers given in AAT03921-31 are used to isolate and/or
CC amplify and/or detect mutations in human Factor-V DNA (AAT03920),
CC esp. in DNA encoding the APC binding and/or cleavage site.
XX
SQ Sequence 6909 BP; 2096 A; 1700 C; 1430 G; 1683 T; 0 other;

Query Match   5.3%; Score 92.8; DB 16; Length 6909;
Best Local Similarity 51.0%; Pred. No. 2.3e-13;

Qy 6386 aatgccaaggacgtgtgaatgctggtcaagccaagccaacaacaatgaagc------ag 6439
Db 1078 tggctgagatcgattgggggaggaaaaaataacaggaattaggagccacagggatct 1137

Qy 6440 tggctgagaattgatctactccaagatcaagaagaagaaacggcaattataacacagggctgc 6499
Db 1138 acacagtcgaacttcaacttttatgttaagagtttgtgatgaacctcaaaaacaataat 1197

Qy 6500 aagtctctgtcctctcctgaaatgtgtaaaagagtataacatccacactacagtgagcaggga 6559
Db 1198 tctaagtgagaaacctataaggaattgtgaataatgagaaagagaagtgttctcaggttaac 1257

Qy 6560 gtggaatggaacacataacaggctgaaatcctccatgtggacaagatcttttgaaggaaat 6619
Db 1258 tctaacttcggaccaagtgcaaaacaattcatcctcccacctcgtggccagatatgtg 1317

Qy 6620 actaataccaaaggacatgtgaagaagaacttttccaaccccccaatcattccaggtttatc 6679
Db 1318 cgggttgtctccccagacaatggccacagaggatagcctgaaggtgagctcattggttgc 1377

Qy 6680 cgtgtcattcctaaaacatgaatcaaagtattgcacttcgcctggaaacctttggctgt 6739
Db 1378 cagatt 1383

Qy 6740 gatatt 6745

RESULT 12

ID AAA60448 standard; cDNA; 6909 BP.
XX
AC AAA60448;
XX
DT 09-OCT-2000 (first entry)
XX
DE Human factor V cDNA sequence SEQ ID NO.26.
XX
DE Factor V; FV; activated protein C; APC; anticoagulant;
KW activated protein C resistant factor V; thrombosis; screening;
KW thrombophilia; ds.
XX
OS Homo sapiens.
XX
PN US6066778-A.
PD 23-MAY-2000.
XX
PF 06-NOV-1996; 96US-0746111.
PR 06-NOV-1996; 96US-0746111.
XX
PA (UNMI) UNIV MICHIGAN.
XX
PI Ginsburg D, Cui J;
XX
DR WPI; 2000-410682/35.
XX
PT New transgenic mice expressing activated protein C resistant factor V
PT and factor V null transgenic mice useful for screening anticoagulants,
PT as models for human thrombophilia and as models for testing in utero
PT gene therapy protocols -

Matches 246; Conservative   0; Mismatches 232; Indels   6; Gaps   1;

Qy   898 aatggttgcagcagatcctttgagttctgaacctgacggggccaaatcagagcttcctcca 957
Db 6280 aatggtgttccaacctccctggttatggaaacggaaagatgaaaacaagcaaatcaca 6339

Qy   958 tggctgagtcgtcaataaggtggagaccaagttcacgttctcctggcccaagccccgactt 1017
Db 6340 gcttctctgttaagaaatctgtatgggagaatctcgggaccctccgtgccgtctcg 6399

Qy 1018 caggacaagcagccatactgggtcttcagggcgcagctagcacacacaccacacagaag 1077
Db 6400 aatgccaaggacgtgtgaatgctggtcaagccaagcaaacaacaatgaagc------ag 6453

Qy 1078 tggctgagaattgggggaggaaaaaataacaggaattaggagccacagggatct 1137
Db 6454 tggctgagaaattgatctactccaagatcaagaagaagaaacggcaattataacaca 6513

Qy 1138 acacagtcgaacttcaacttttatgttaagagtttgtgatgaacctcaaaaacaataat 1197
Db 6514 aagtctctgtcctctcctgaaatgtgtaaaagagtataacatccacactacagtgagcaggga 6573

Qy 1198 tctaagtgagaaacctataaggaattgtgaataatgagaaagagaagtgttctcaggttaac 1257
Db 6574 gtggaatggaacacataacaggctgaaatcctccatgtggacaagatcttttgaaggaaat 6633

Qy 1258 tctaacttcggaccaagtgcaaaacaattcatcctcccacctcgtggccagatatgtg 1317
Db 6634 actaataccaaaggacatgtgaagaagaacttttccaaccccccaatcattccaggtttatc 6693

Qy 1318 cgggttgtctccccagacaatggccacagaggatagcctgaaggtgagctcattggttgc 1377
Db 6694 cgtgtcattcctaaaacatgaatcaaagtattgcacttcgcctggaaacctttggctgt 6753

Qy 1378 cagatt 1383
Db 6754 gatatt 6759

RESULT 7

```
Db   474  rssnefpaqteqttaspdirnttvtpnvtkdvalaavlvpvlvmvltllilllvcawhvr 533
Qy   528  -RKKKKKGS---PY 537
          ::|| | ||
Db   534  nrkkkefgtydlpy 547
```

```
ID   AAY70539 standard; protein; 669 AA.
XX
AC   AAY70539;
XX
DT   04-JUL-2000 (first entry)
XX
DE   Human Factor 8 Homologue.
XX
KW   Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW   cerebroprotective; therapeutic; coagulation related disorder;
KW   haemophilia; stroke; screening.
XX
OS   Homo sapiens.
XX
PN   WO200012532-A1.
XX
PD   09-MAR-2000.
XX
PF   20-AUG-1999;
XX
PR   31-AUG-1998; 98US-0098521.
XX
PA   (ELIL ) LILLY & CO ELI.
PI   Rosteck PRJ, Su W, Li XM;
XX
DR   WPI; 2000-256580/22.
DR   N-PSDB; AA251872.
XX
PT   Factor 8 homolog polypeptides and nucleic acids encoding them, for
PT   treating coagulation related disorders such as haemophilia and stroke
XX
PS   Claim 3; Page 64-66; 68pp; English
XX
CC
CC   The present sequence is a human Factor 8 homologue (F8H),
CC   a coagulation cofactor which is selectively expressed in
CC   haematopoietic, heart and reproductive tissues. It has haemostatic and
CC   cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC   and is useful as a therapeutic for treating coagulation related diseases
CC   such as haemophilia and stroke. The nucleic acid is useful as
CC   hybridisation probe and amplification primer for detecting deficiencies
CC   in the level of F8H mRNA, for screening F8H gene mutations and for
CC   monitoring regulation of gene expression. Fragments of the nucleic acid
CC   are also useful as diagnostic probes and primers, and can be used in
CC   screening methods such as those using DNA chips. The present sequence is
CC   also useful as a target to screen therapeutically useful modulators
CC   of the F8H.
XX
SQ   Sequence 669 AA;
```

```
Query Match     35.8%;  Score 786.5;  DB 21;  Length 669;
Best Local Similarity  35.8%;
Matches  190;  Conservative  94;  Mismatches 164;  Indels  81;  Gaps  16;

Qy   124  GKRILRLGDLDIE-SQTCASDYLLPTS----SSDQYGPYCG-SMTVKKELLNTSEVTV 177
          | |: | ||:|  ||  :    |    |       |  || ::  |||| |:|
Db   178  RFESGSHISGRGFLLTYASSDHPDLITCLERASHYLKTEYSKFCPAGCRDVGDISGNMV 237

Qy   2    gevrrikfgdfdlededschfnylriylpngisvartelskyeglqlmqnhleskqneitl 61

Qy   178  RFESGSHISGRGFLLTYASSDHPDLITCLERASHYLKTEYSKFCPAGCRDVGDISGNMV 237

Qy   62   lfmsglhvsgrrgflasysvldkqdlitcldteasnfiepefskycppagclpfaesgtlp 121

Db   238  DGYRDTSLLCKAAIHAGIITADELGGQISVLQRRGISRYEGILANGVLSRDGSLSDRAFLF 297
```

RESULT 8

```
Db   122  hgyrdssplcmagyhagvvsntlggqisvvlskdjpyyesslannvtsvvghlstslff 181
Qy   298  TSNGCSRSLSFE----PDGQIRASSSWOSVNBSGCDQVHWSPGGQARLQDOGPSNASGDSSN 353
Qy   182  Resgcygtlgmesgvladpqltassvlewtdhtggenswapkkarlkkppppwaafatde 241
Qy   354  NIHQKREWLEIDLGEKKTIYGIRTYGSTQSNFNFYVKSPVNNFKNNNSKWKTYKGIVNNE 413
Db   354  ---ywlqidlnkekkitgiiittgstmvehnyyvsayrilysddgqkwtvyrepgvegd 297
Qy   414  KVFQGNESNFRDPVQONNFIPPIVARYVRVVPQTWHQRIALKVELIGCQ------ITQ- 463
Db   414  kifqgnkdVnqdvrnnflpplilarflrvnptcqwqkkiamkmelilgcqfipkgfrppklcq 357
Qy   464  ---GNDILVWRKTSQSTVSYSTKEDEVITRPI---------PSEET-STGINITT- 505
Db   358  pppprnsnd---lkottappkiakgqrapkfcqplqprssnefpaqteqttaspdirnttv 413
Qy   506  ---VAIPLVLLVWVFFAGMGIFAAF------RKKKKKGS---PYGSAEAGKTDC 547
Db   414  tpnvtkdvalaavlvpvlvmvltllilllvcawhvrnrkkkefgtydlpyw----dragw 469
Qy   548  WKQIKIPF-----------ARDGSAEFTISYDNEKEMTOKLDLIITSDMA 585
Db   470  wkgmkqflpakavdheetpvryysse-vnhlsprevtlvlqadsaeya 516
```

```
ID   AAB24216 standard; protein; 889 AA.
XX
AC   AAB24216;
XX
DT   06-FEB-2001 (first entry)
XX
DE   Soluble neuropillin.
XX
KW   Mouse; soluble neuropillin.
XX
OS   Mus sp.
OS   Synthetic.
XX
PN   JP200236879-A.
XX
PD   05-SEP-2000.
XX
PF   17-FEB-1999;  99JP-0038920.
XX
PR   17-FEB-1999;  99JP-0038920.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI; 2000-658506/64.
DR   N-PSDB; AAC60787.
XX
PT   Soluble neuropillin having a deleted cell membrane-penetrating region,
PT   used to diagnose human disease conditions
XX
PS   Claim 5; Page 13-17; 18pp; Japanese.
XX
CC
CC   The present invention describes a recombinant soluble neuropillin (I)
CC   comprising an amino acid sequence in which the cell membrane-penetrating
CC   region and/or the intracellular region is deleted from a mouse-derived
CC   neuropillin. The soluble neuropillin can be used for the elucidation of
CC   human bioreactions and disease conditions. The present sequence
CC   represents a specifically claimed soluble neuropillin protein
CC   sequence from the present invention.
XX
SQ   Sequence 889 AA;
```

```
Query Match     11.6%;  Score 353.5;  DB 21;  Length 889;
```

Best Local Similarity 23.1%; Pred. No. 2.3e-22;
Matches 120; Conservative 73; Mismatches 189; Indels 137; Gaps 17;

```
Qy   49 VPGARGGGALARNACRGILALLLAVSAPLRLQAEELGD-GCGHLVTYQDSGTMYSKNPG 107
            |||.||| |: |: :||  | :||
Db   15 vpgstgdaa------qparavrslqdppcqgrpsrkdagyitspgypq 57

Qy  108 TYPNHTVCEKRIITVPK-GKRLILRLG-DLDIESSQTCASDYLLF----TSSSDQYGPYCGS 161
            |:|| :|  :|||  ||   ||  || :|| |
Db   58 dypshqacewivyapepnqkivinfnphfeiekhackydfieirdgsdesadllgkhcgn 117

Qy  162 MTVPKELLLNTSEVTVRFESSGSHISGRGFLLTY------ASSD 198
            | |: :||     :| ||    |
Db  118 i-apptilssgvsivlykfsdyarqgagfslryelfktgsedcsknftspngtiesppfp 176

Qy  199 -HPDLITC------LERASHYLKTEYSKF----------CPAGCRDVAGDIS--GNMV 237
            |  ||     |    |:|                  ||  | | :|
Db  177 ekyphidctftilakprmeilliqflftfdlehdplqvgegdckydwldidviphvgpli 236

Qy  238 DGTRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRY----------EG 277
            :|   :  | |     |    |   :|
Db  237 gkycgtktpskirsstglislfhtdmavakdgfsaryylhbgeppenfqcnvplgmesg 296

Qy  278 ILANGVLSRDGSLSDKRFLPTSNGCSRSLSFEPDGQIRASSSWQSVNESGDQVHWSPGQA 337
            | ||       |                             :|
Db  297 rlaneqisasstfsdgr----------------------wtpqgs 319

Qy  338 RLQDQGPSWASGDSSNNHKPRPWLEIDLGEKKKTTGIRTTG---STQSNFNF 395
            || :    |    |      |   |   ||  | |:|   :|| ||:
Db  320 rlhgddnqwtpnldsn-------keylqvdlrfltmltaicqaisretqkyyvksykiev 375

Qy  396 KNNNSKWKTYKGIYNNERKVFQGNSNFRDPVQONFIPPIVARYVRVVPQTWHQRIALKVE 455
            .|  . | |  | :|      | |   :||  | | ||:|:| |   |
Db  376 stngedwmvyrhgknh--kifqanndatevnlkhlmphltrfiirpqtwhighlalrle 433

Qy  456 LIGCQITCQGNDSLVWRKTS--QSTSVSTKKEDETIRP 491
            | || :|  : :|                    :|
Db  434 lfgcrvtdapcsnmlgmlsglladtqisasstreylwsp 472
```

SQ   Sequence 925 AA;

RESULT 9
AAW96308
ID   AAW96308 standard; Protein; 925 AA.
AC   AAW96308
DT   28-JUN-1999 (first entry)
DE   Neuropilin-2.
PA   Ginty DD, Kolodkin AL;
PI   (UIJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
PN   WO9904263-A1;
PD   28-JAN-1999;
PF   17-JUL-1998;
PR   9US-0052762.
PT   New isolated semaphorin receptor, neuropilin-2 - used to develop

PT   products for the diagnosis and treatment of neurological,
PT   immunological, oncological and viral diseases

Claim 13; Page 77-79; 88pp; English.

CC   The neuropilins are type I transmembrane proteins and act as
CC   semaphorin III (Sema III) receptors. The semaphorins have been
CC   shown to function in repulsive axon guidance. Sema III is a
CC   secreted protein that in vitro causes neuronal growth cone
CC   collapse and chemorepulsion on neurites and is required in vivo for
CC   correct sensory afferent innervation and other aspects of
CC   development. Agents which inhibit or enhance the interaction of a
CC   semaphorin and a neuropilin can be potent modulators of nerve cell
CC   growth, immune responsiveness, and viral pathogenesis, and can be
CC   used in the treatment and diagnosis of neurological disease,
CC   neuro-regeneration, immune modulation including hypersensitivity and
CC   graft-rejection, and diagnosis and treatment of viral and oncological
CC   infection/diseases. The neuropilins, neuropilin-encoding nucleic
CC   acids, and unique portions also are useful in screening chemical
CC   libraries for regulators of semaphorin-mediated cell activity, and in
CC   genetic mapping as probes for related genes, as diagnostic reagents
CC   for genetic, neurological, immunological and oncological disease.

SQ   Sequence 925 AA;

Query Match   11.5%; Score 351.5; DB 20; Length 925;
Best Local Similarity 24.4%; Pred. No. 3.7e-22;
Matches 114; Conservative 76; Mismatches 181; Indels 97; Gaps 17;

```
Qy   88 CGHLVTYQDSGTMYSKNPGTYPNHTVCEKRIITVPK-GKRLILRLG-DLDIESOTCASDY 145
            || :||   |   :|| :|  :|||  ||   ||  || :|| |
Db   28 cggrlnskdagyitspgypqdypshqacewivyapepnqkivlnfnphfeiekhackydf 87

Qy  146 LLF----TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESSGSHISGRGFLLTY----AS 196
            | |         | ||   | |: :||     :| ||    |
Db   88 leirdgsdesadllgkhcgni-apptilssgvsivlykfsdyarqgagfslryelfktgs 146

Qy  197 SD----HPDLITC------LERASHYLKTEYSKF---------- 220
            :|    |  ||      |    |:|
Db  147 edcsknftspngtiesppfpekyphidctftilakprmeilliqflftfdlehdplqvgeg 206

Qy  221 -CPAGCRDVAGDIS--GNMVDGTRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRY-- 275
            ||  | | :|    :|   :|   :  | |     |    |   :|
Db  207 dckydwldidviphvgpligkycgtktpskirsstglislfhtdmavakdgfsaryyl 266

Qy  276 ------EGILANGVLSRD-GSLSDKRFLPTSNGCSRSLSFEPDGQIRASSSWQSVNESGD 328
            | ||       |
Db  267 vhbgeppenfqcnaplgmesgrlane----------------qisasstf------s 300

Qy  329 QVHWSPGQARLQDQGPSWASGDSSNNHKPRPWLEIDLGEKKKTTGIRTTG---STQSNFNF 386
            : || :    |    |      |   |   ||  | |:|   :|| ||:
Db  301 dgrwtpqgsrlhgddnqwtpnvdsn----keylqvdlrfltmltaicqaisretqkyy 356

Qy  387 YVKSFYNNFKNNNSKWKTYKGIYNNERKVFQGNSNFRDPVQONFIPPIVARYVRVVPQTW 446
            |||.  | |  . | |  | :|      | |   :||  | | ||:|:| |
Db  357 yvksykievstngedwmvyrhgknh--kvfqanndaicevlnkhtplltrfiirpqctw 414

Qy  447 HQRIALKVELIGCQITCQGNDSLVWRKTS--QSTSVSTKKEDETIRP 491
            | |:|| :| || :|  : :|                    :|
Db  415 hiqlalrlelfgcrvtdapcsnmlgmlsglladtqisasstreylwsp 462
```

RESULT 10
AAB64627
ID   AAB64627 standard; Protein; 439 AA.
AC   AAB64627
DT   22-MAR-2001 (first entry)
DE   Human secreted protein BLAST search protein SEQ ID NO: 137.

SEQ ID NO:3
AC NO: AAA96736, Database: N-Geneseq-032802

```
CC   perturbing the normal function of NHP in the body. Nucleotide constructs
CC   encoding NHP products are used to genetically engineer host cells to
CC   express such products in vivo. These host cells allow for the
CC   identification of compounds that bind to NHP receptors or trigger
CC   NHP-mediated pathways.
XX
SQ   Sequence 1768 BP;
```

```
Query Match             86.48%;  Score 1522;  DB 22;  Length 1768;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1522; Conservative   0; Mismatches   0; Indels   0; Gaps   0;
```

```
Qy    240 gggcggaggagctggggtgatggcctgtgagacacctagtgacttatcaggatagtgcacat 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 gggcggaggagctggggtgatggcctgtgagacacctagtgacttatcaggatagtgcacat 60

Qy    300 gacatctaagaattatccggggaggacctcacccaatacacactgttgcgaaaagacaattac 359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 gacatctaagaattatccggggaggacctcacccaatacacactgttgcgaaaagacaattac 120

Qy    360 agtaccaaaggggaaagagactycattctgaggttggggaggttcggatacgaatcccaagc 419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 agtaccaaaggggaaagagactycattctgaggttggggaggttcggatacgaatcccaagc 180

Qy    420 ctgtcgttctgtactatctctctcaccagctcctctcagaactacatatgtccatactgtgg 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ctgtcgttctgtactatctctctcaccagctcctctcagaactacatatgtccatactgtgg 240

Qy    480 aagtagtactgttccaaagaactcttgtgtgaacacaagtgaagtaaccgttccggcttga 539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aagtagtactgttccaaagaactcttgtgtgaacacaagtgaagtaaccgttccggcttga 300

Qy    540 gagtgagatcaccacattctggccgggggttttcttttgtgacctatgcggaggaccatcc 599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gagtgagatcaccacattctggccgggggttttcttttgtgacctatgcggaggaccatcc 360

Qy    600 agattaataacactgtttggaacgagcagctagctattattgaagacagaatacagtgaaatt 659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 agattaataacactgtttggaacgagcagctagctattattgaagacagaatacagtgaaatt 420

Qy    660 ctgcccagctcggttctagacagcctagcagtacagaatcagcgaatt 719
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ctgcccagctcggttctagacagcctagcagtacagaatcagcgaatt 480

Qy    720 tagagtatcctcttcattgtcaaagctgccactctgtgcaggaataatgtcgatgaact 779
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 tagagtatcctcttcattgtcaaagctgccactctgtgcaggaataatgtcgatgaact 540

Qy    780 agttgccagatcacagtctgcttcaggcgacaaagggatacaaggatttctggc 839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 agttgccagatcacagtctgcttcaggcgacaaagggatacaaggatttctggc 600

Qy    840 caattgtgttcttcgaggacgtgttccctgtcagacaacgatttctgttacctccaa 899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 caattgtgttcttcgaggacgtgttccctgtcagacaacgatttctgttacctccaa 660

Qy    900 tggttgcagcagatcccttggagtttgaacctgacgggcaaatcaggagcttcttcctcatg 959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 tggttgcagcagatcccttggagtttgaacctgacgggcaaatcaggagcttcttcctcatg 720

Qy    960 gcagtcggtcaatgagagtgggagaccaagtcaactgggccaagcccgacctca 1019
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gcagtcggtcaatgagagtgggagaccaagtcaactgggccaagcccgacctca 780

Qy   1020 ggaccaaggcgccatcaatgggcgcttcggggcggacaagtgccagtcaacacacacacagagtc 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ggaccaaggcgccatcaatgggcgcttcggggcggacaagtgccagtcaacacacacacagagtc 840

Qy   1080 ggctggaagatcgtattcgggggaagaaaaaataacaaggagaaataacaaggaatatac 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ggctggaagatcgtattcgggggaagaaaaaataacaaggagaaataacaaggaatatac 900
```

```
Qy   1140 acagtcgaacttcaacttttatgtcttaagagtttttgtgatgaagtgtgaacttcaaaaacaataatc 1199
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 acagtcgaacttcaacttttatgtcttaagagtttttgtgatgaagtgtgaacttcaaaaacaataatc 960

Qy   1200 taagtggaagacctataaaggaagtaatttgtgaataatgaagaaaaggttgaatgaactc 1259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 caagtggaagacctataaaggaagtaatttgtgaataatgaagaaaaggttgaatgaactc 1020

Qy   1260 taacttcgggaccagtcgcaaacaatttcatcctcccagatactgtggccagaatgctgcg 1319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 taacttcgggaccagtcgcaaacaatttcatcctcccagatactgtggccagaatgctgcg 1080

Qy   1320 ggttctccccagacatggcaccagcagaggatagccttgaaaggtggagtcaatggttgcca 1379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ggttctccccagacatggcaccagcagaggatagccttgaaaggtggagtcaatggttgcca 1140

Qy   1380 gattacacaaaggtaatgattcattggtgtcgcgaagacacaagagacacagttc 1439
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 gattacacaaaggtaatgattcattggtgtcgcgaagacacaagagacacagttc 1200

Qy   1440 aactaagaagatgtgaatactgaagaaataacatcacaaccacagtgttc 1499
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 aactaagaagatgtgaatactgaagaaataacatcacaaccacagtgttc 1260

Qy   1500 aataaacattacagacggtggctattcccttgtctcctggtcttctctgttgttctgctgg 1559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 aataaacattacagacggtggctattcccttgtctcctggtcttctctgttgttctgctgg 1320

Qy   1560 aatgggatcttgctccctttggaaagaagaagaagagtccgtatggatcagc 1619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 aatgggatcttgctccctttggaaagaagaagaagagtccgtatggatcagc 1380

Qy   1620 ggaggctcagaaaacagacggttgggaggaagattaaatatcccttgccagacatcagtc 1679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ggaggctcagaaaacagacggttgggaggaagattaaatatcccttgccagacatcagtc 1440

Qy   1680 agctgagtttacccctaccaatgaaaagacagacgagacacaaaagttagatccat 1739
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 agctgagtttacccctaccaatgaaaagacagacgagacacaaaagttagatccat 1500

Qy   1740 cacaagtgatatggcaagttaa 1741
          ||||||||||||||||||||||
Db   1501 cacaagtgatatggcaagttaa 1522
```

```
RESULT 5
AAA96736
ID   AAA96736 standard; DNA; 1871 BP.
XX
AC   AAA96736;
XX
DT   19-FEB-2001 (first entry)
XX
DE   Polynucleotide isolated from lymph node stromal cells of fsn -/- mice.
XX
PN   WO200058463-A1.
PD   05-OCT-2000.
XX
PF   18-FEB-2000; 2000WO-NZ00015.
XX
KW   Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW   immune system disorder; cancer; viral disorder; HIV infection;
KW   blood vessel growth; tumour necrosis factor disorder; arthritis;
KW   inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW   cardiac failure; ss.
OS   Mus sp.
XX
FH   Key            Location/Qualifiers
FT   CDS            64..1575
FT                  /tag= a
XX
```

PR   25-MAR-1999;
PR   26-AUG-1999;
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
PI   Murison JG;
XX
PR   9905-0276268.
PR   9905-0383586.
XX
WPI; 2000-664924/64.
DR   P-PSDB; AAB19126.
XX
PS   Claim 4; Page 60-61; 75pp; English.
XX
CC   The present sequence represents a polynucleotide sequence which is
CC   isolated from lymph node stromal cells of fsn -/-. lymph node stromal cells,
CC   and their polypeptides are useful for treating inflammatory
CC   and tumour necrosis factor-mediated disorders, cancer and viral
CC   disorders.
CC   Polypeptide expressed in mammalian fsn -/-. lymph node stromal cells,
CC   useful for modulating growth of blood cells, for treating inflammatory
CC   and tumour necrosis factor-mediated disorders, cancer and viral
CC   disorders.
CC   The polynucleotides and their polypeptides are useful for treating an
CC   inflammatory disorder, disorder of immune system and cancer selected
CC   from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
CC   viral disorder, in particular HIV infection and for modulating the
CC   growth of blood vessels. The polypeptides are useful for treating a
CC   tumour necrosis factor (TNF) mediated disorder, such as those selected
CC   from arthritis, inflammatory bowel disease and cardiac failure and a
CC   fibroblast growth factor-mediated disorder. It is also useful in assays
CC   to determine biological activity, to raise antibodies, to isolate
CC   corresponding ligands or receptors, to quantify levels of protein or
CC   cognate corresponding ligand or receptors, as antiinflammatory agents,
CC   and in compositions for the treatment of skin, connective tissue and
CC   immune system diseases. The polynucleotide is useful as marker for
CC   tissue, as a chromosome marker or tags in the identification of a
CC   genetic disorder.
XX
SQ   Sequence 1871 BP; 485 A; 471 C; 484 G; 431 T; 0 other;

Query Match   30.6%; Score 539.2; DB 21; Length 1871;
Best Local Similarity   82.8%;
Matches 616; Conservative   0; Mismatches 128; Indels   0; Gaps   0;

Pred: No.1.7e-128;

```
Qy     159 cggcggcggcggcgcactggcgcggcgtgccgggggggccctcggcttgtccgccggt 218
           |||| || | || |||| |||||||||||  || |||| |||||||||||||||||
Db      54 cggcggaccatggggactggtcgggccgagtcctggcgctggctgttccgccgt 113

Qy     219 ctccgcccccgctccggctgcaggcggaggaggctgggctgatgggtggacacctagtgac 278
           |||  |||||||| |||||||||||||||||||||||| ||||| |||||||||||||
Db     114 gtgtctccgctccggtcgacggcggaggaggctgggtgatgggtgggacacagtgac 173

Qy     279 ttatcaggatagtgcacaatgacatctaagaattatcccgggacctcaccccatcacac 338
           ||||| ||||| |||||||||||||||||||||||||| || |||||||| ||||||
Db     174 ctctcaggacacgtcgcacaatgacatctaagaattatccaggagacttaccccaattacac 233

Qy     339 tgtttgcgaaagacaattacagtacaaagggaaaagaactgagcttgtccagcgtccggtgggaga 398
           |||  |||| |||| || | ||| |||||| ||||| ||| ||||||  ||||||||||
Db     234 tgtgtcgtaaaagatcatcacacgtcccaaagggaagagacttatcctgaagttgggaga 293

Qy     399 tttggatatcagaatcccagacgctgcttccgactatctctctccacacgctcttcaga 458
           ||| || |||| ||||||  || | |||||||||||||| | |||||||||| ||||||
Db     294 ttttgaacattgagtccaagaccgcctctcgactatctcctcttcagacgtgcaaacag 353

Qy     459 tcacatatggtccacactgggaagaagtataacgttcccaaagaaaagaactcttgtgacacacaag 518
           ||| ||||||||| | |||| ||||| |  ||| ||||||||||||||| ||  ||||||| || ||
Db     354 tcagtatggtccatatggtggagtgggctggttgttcccaaagaaaagctccggctgtgacactcaag 413

Qy     519 tgaagtaaacctgccgtcttgaagtggatccacacattcctggccgggggttttttctgac 578
           | ||||  | |||||| |||||||||||||||||||||||||||||||||| || ||||||
Db     414 cgaagtgacatgctcctccaagagtgggatccacacattcctggccgggggcttccgctgac 473

Qy     579 ctaatgcgagcagcgaccatccagattaataacatgtttgaacgactagccaattatt 638
           ||||| | || |||||||||||||||| ||||| ||||||| || || ||||||||||
Db     474 ctacgccagcagtgaccatccagattaataacctgtttgggaacgacgagcggcagccattatt 533

Qy     639 gaagcagagaatacagcagaaattctgcccagtctggttgttgagagacctgacagagagacattgc 698
           | || ||| | ||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Db     534 cgaggaaaatacagcaaaattctgcccagtctggttgttgagagacctgacagagagacattgc 593

Qy     699 tgggaatatgtaggatgatatatgaagataaccctcttttattgtgcaaagctgccatcgc 758
           ||||||| | | ||||  |  | ||||| ||||||||||||||||||||||||||||
Db     594 tgggaatacaaaagatgttacagagataacctcttttattgtgcaaagctgcca 653

Qy     759 aggaataattgctgatgaactaggtgtggccaagatcagtgtgtccagcgaaaggtgatcag 818
           ||| || || |||||||||||||||||||||| | | ||   ||| || |||||| || ||
Db     654 agggatcatcacagatgaactaggtgtggccatcacaacttgcttcagagcaaaaggataag 713

Qy     819 tcggatatgaaggggattctggccaaggatggtggttcttcagccgcagaaaggggatcaca 878
           || |||| ||||| ||| |||||||| || || ||||| ||| |||| |||  || ||||
Db     714 tcaactaatgaaggggactcctggccaatggctgcgctccccgtcaaggagctggttcttttcggaaa 773

Qy     879 gcgattctcgttacctccaatgg 902
           ||||||||| || | |||||
Db     774 gcgattctcttttacaacccagg 797
```

Result   6
AAH34855
ID   AAH34855 standard; cDNA; 2428 BP.
XX
AC   AAH34855;
XX
DT   08-SEP-2001 (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:1937.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200129920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US026524.
XX
PR   29-SEP-1999; 9905-0157137.
PR   03-NOV-1999; 9905-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
WPI; 2001-235357/24.
DR   P-PSDB; AAG75450.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 3444-3445; 9803pp; English.
XX
CC   AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the patients own production of P.
CC   Additionally, N may be used to supplement the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

OM protein - protein search, using sw model

Run on: July 9, 2002, 01:13:23 ; Search time 69.09 Seconds

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Title: (without alignments)
Perfect score: 3052   815.000 Million cell updates/sec
Sequence: 1 MGFGAGQRLRPVPAPRSSAE.........YDNEKEMTVQKLDLITSDMAG 586

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 283118 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283118

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	331	10.8	2211	1	KFBO5	coagulation factor
2	330	10.8	927	1	JC0948	As antigen precurs
3	318	10.5	2224	1	KFHU5	coagulation factor
4	318	10.4	2183	1	T42764	coagulation factor
5	299	9.8	218	2	A47285	milk fat globule p
6	297.5	9.8	427	2	JC4915	milk fat globule p
7	285.5	9.4	216	2	A44258	ags protein precur
8	285.5	9.4	2351	1	EZHU	factor VIII-associ
9	285.5	9.4	463	2	A36479	coagulation factor
10	276.5	9.1	2319	2	A47004	milk fat globule m
11	271.5	8.9	2133	2	T42763	coagulation factor
12	268	8.8	401	2	S65138	coagulation factor
13	268	8.8	427	2	S74211	glycoprotein antig
14	259.5	8.6	409	2	T11743	PAS-6/7 protein pr
15	217	7.1	845	2	JC5256	pp47 protein - pig
16	216.5	7.1	719	2	S51739	adipocyte transcri
17	192.5	6.3	1283	2	T13799	transcription repr
18	187	6.1	3133	2	S52093	neurexin IV - frul
19	186.5	6.1	666	2	A59271	hemocytin - silkwo
20	174	5.7	1524	2	T30337	Ra-reactive factor
21	172.5	5.7	319	2	151569	polyprotein - Afri
22	168	5.5	3623	2	T09456	UVS.2 protein - Af
23	165.5	5.4	986	2	B58788	intrinsic factor-B
24	165	5.4	730	1	BMHUI	procollagen C-endo
25	165	5.4	823	1	A58788	procollagen C-endo
26	163.5	5.4	449	2	A55362	procollagen C-endo
27	163.5	5.3	819	2	I48059	procollagen I C-pr
28	160.5	5.3	694	2	JC6554	tyro 10 receptor k
29	158	5.2	991	2	I49540	complement subcomp
30	156	5.1	707	2	JC2218	procollagen C-endo
31	156	5.1	3623	2	T08618	intrinsic factor-B
32	155.5	5.1	855	2	S42621	protein-tyrosine k
33	153	5.0	402	2	JH0403	procollagen I C-pr
34	149.5	4.9	695	1	S05008	complement subcomp
35	146.5	4.8	1019	2	A38738	coagulation factor
36	142.5	4.7	1385	2	T14158	neurexin IV - mous
37	141	4.6	688	1	C1HUS	complement subcomp
38	141	4.6	2083	2	T42721	CRP-ductin-alpha r
39	140.5	4.6	1381	2	T31083	paramodrin rat
40	137	4.5	705	1	C1HURA	Ra-reactive factor
41	136.5	4.5	599	1	T15763	complement subcomp
42	135	4.4	737	1	T15615	hypothetical prote
43	133.5	4.4	1594	2	T30549	hensin - rabbit
44	133	4.4	855	2	JC7731	membrane-bound arg
45	133	4.4	855	2	JC7775	membrane type-seri

ALIGNMENTS

RESULT 1

KFBO5
coagulation factor V precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999
C;Accession: A42580; A36497
R;Guinto, E.R.; Esmon, C.T.; Mann, K.G.; MacGillivray, R.T.
J. Biol. Chem. 267, 2971-2978, 1992
A;Title: The complete cDNA sequence of bovine coagulation factor V.
A;Reference number: A42580; MUID:92147638
A;Accession: A42580
A;Molecule type: mRNA
A;Residues: 1-2211 <CU1>
A;Cross-references: GB:M81440; NID:g1630037; PIDN:AAA30512.1; PID:g1630038
R;Kalafatis, M.; Jenny, R.J.; Mann, K.G.
J. Biol. Chem. 265, 21580-21589, 1990
A;Title: Identification and characterization of a phospholipid-binding site of bovine
A;Reference number: A36497; MUID:91072154
A;Accession: A36497
A;Molecule type: protein
A;Residues: 1566-1570,'X',1572-1581,'X',1583-1584;1673-1676,'X',1678-1679,'X',1681,'X
R;Xue, J.; Kalafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.
Biochemistry 33, 13109-13116, 1994
A;Title: Determination of the disulfide bridges in factor Va heavy chain.
A;Reference number: A55979; MUID:95034740
A;Contents: annotation
A;Note: 566-Cys and 617-Cys were shown to have free sulfhydryls
A;Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
C;Function:
A;Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa p
A;Pathway: blood coagulation
C;Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidas
C;Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plas
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-2211/Product: coagulation factor V #status predicted <MAT>
F:29-741/Product: coagulation factor Va heavy chain #status predicted <VAH>
F:29-345/Domain: A1 <DA1>
F:33-229/Domain: ferroxidase repeat homology <FO1>
F:346-695/Domain: A2 <DA2>
F:351-688/Domain: ferroxidase repeat homology <FO2>
F:696-1564/Domain: B <DOB>
F:1175-1437/Region: 9-residue repeats (O-X-T/N-L-S-P-D-L-S)
F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>
F:1572-1892/Domain: A3 <DA3>
F:1572-1892/Domain: ferroxidase repeat homology <FO3>
F:1654-1752/Region: phospholipid binding #status predicted
F:1893-2051/Domain: C1 <DC1>
F:1893-2048/Domain: discoidin I amino-terminal homology <DN1>
F:2052-2211/Domain: C2 <DC2>
F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>
F:167-193,248-329,499-525/Disulfide bonds: #status experimental

```
Qy   66 LLALLLAYSAPLRLQAEELGDGCGHLVTYQDSGTMTSKNYPGTYPNHTVCEKIITVPKG-  124
        |  :  |            | |    :  |  ::    |  |  |  :|| ::|:|    |
Db    5 LLSCCCWLLCSLRSSWASRNDKCGDTIKITSPSYLTSAGTPHSTYPSQORCEMLIQAPEHY  64
```

RESULT 2

```
Qy  209 ASHYLKTEY-SKFCPA------------GCRDVAGDISGNNV----DGYRDTSLLCKAAIHAGIIADE  259
        ||  |:::  |:|  |             | : | :  | |     ||  | ||||||  ||| | |
Db 1968 AKHYLKPYITEFCVAYSLDRKNWRIFKGNSTRNVWPFGNSDAS?IKENQIDPPVVA---  2025
```

```
Qy  260 LGCGOISVLQRKGISRYEGIILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE-------PDGQIR  315
        |||   |  : | | |   |   ::   ||   :| : | | |    ::       | | |
Db 2026 RVIRISPFGSYNKPALRLELGGCG-EVNGCGSTPLGMESGKIENKOIT  2070
```

```
Qy  316 ASS------SWQSVNESGCDVIHWSPGQARLQDGG--PSWASGDSSNNHKPREWLEIDLGRKKK*  370
        |            |    :  |:| |  |    | |    | |    |:   :   |   |
Db 2071 ASSPKKSWWG-------NTWEPFLARLNAQGRVNAWQAKANNNN---QWLQIDLLKIKR  2119
```

```
Qy  371 ITGIRTTGSTQSNEFYVKSEFYVMNFKNNNSKWKTTKGIVNNEEKVFQGNSNFRDPVYNNF  430
        | |  |  ||  ||  |:|||  |:| :| | | |||| : |||||| |:| ||  ||
Db 2120 ITAIVTQGGCKSLSSEMYVKSTYIHYSDGGTDKNKPTRERSSMVQIKIFEGNNNVRGHVKNFF  2179
```

```
Qy  431 IPPIVARYVRVVPQTWHQRIALKVELIGCGCI  461
        | | || ||  ||| || |||||||||| | |
Db 2180 NPFIISRFRIIPKTWNQSTALRLELFGCDM  2210
```

Query Match   10.8%;   Score 331;   DB 1:   Length 2211;
Best Local Similarity   32.5%;   Pred.No.   6.1e-16;
Matches   88;   Conservative   48;   Mismatches   89;   Indels   46;   Gaps   10:

A; Molecule type: mRNA
A; Accession: JH0466 <TXM>
A; Residues: 1-927 <CRI>
A; Reference number: JH0466; MUID:91337458
A; Title: The A5 antigen, a candidate for the neuronal recognition molecule, has homolog
A; R:Takagi, S.; Hirata, T.; Agata, K.; Mochii, M.; Eguchi, G.; Fujisawa, H.
A; Neuron 7, 295-307, 1991
C; Accession: JH0466; JQ0948
C; Date: 10-Sep-1999 #sequence_revision 10-Sep-1999
C; Note: this protein has motifs homologous to complement components C1r and C1s and to
C; Comment: This protein is a neuronal cell surface molecule involved in the neuronal reg
C; Superfamily: Xenopus A5 antigen; glycoprotein; transmembrane protein
C; Keywords: duplication; signal sequence #status predicted <SIG>
C; -1-21/Domain: signal sequence #status predicted <SIG>
F; 22-927/Product: A5 antigen #status predicted <A5A>
F; 27-138/Domain: C1r/C1s repeat homology <C1R1>
F; 147-262/Domain: C1r/C1s repeat homology <C1R2>
F; 274-424/Domain: discoidin I amino-terminal homology <CNI>
F; 430-584/Domain: discoidin I amino-terminal homology <DN2>
F; 644-812/Domain: MAM homology <MAM>
F; 861-883/Domain: transmembrane #status predicted <TMM>
F; 150,261,300,523,844/Binding site: carbohydrate (Asn) (covalent) #status predicted

A; Species: Xenopus laevis (African clawed frog)
A; Molecule type: mRNA
A5 antigen precursor - African clawed frog
JQ0948   2

```
Qy  125 KRLLIRLG-DLDIESQTCASDYLLF------TSSSDQYGPYCGSMYVPKELLLNTSEVTVRF  179
        |  | | |           :||        ::::| |   |  : :  | | |:| : ||
Db  125 ORIMINENPHPDLEDRECKYDYEVIDGDNANGCGLLGKYCGKI-APSPLVSTGPSIPIRF  183
```

RESULT 3

```
Qy   65 ESGSHISGRGFILTY-----------------------ASSD---------------HPDLITC---  123
        |||   | :| :|                          |||               | | |
Db   65 LGMESGEIHPDQISVSSQY--------SMNNSAERSRLNYVENCWTPGEDT----VKEMI  205
```

```
Qy  180 LERASHYLKTEYSKFCPAG--CR-DVAGDISGNMVDGYRDTSLLCKAAIHAGIIADE  259
        | | |     :||                | :  | | ||  : | ||||  ||| | |
Db  206 QVDLENLRFVSGIGTQGAIGSKEFKKVKYKVDLISSNGEDWITLKD--GNKIHVPTGN  384
```

```
Qy  260 LGCGOISVLQRKGISRYEGILA-------NGVLSRG-------SLSDKRFLFTSNGCSRS  305
        |||  |     | | | |   |            |       | | |  | |  ||
Db  327 SNFRDPVQNNFIPIVARYVRVVPQTWHQRIALKVELIGCGQIT  462
```

Query Match   10.8%;   Score 330;   DB 1:   Length 927;
Best Local Similarity   24.2%;   Pred.No.   2e-16;
Matches   112;   Conservative   74;   Mismatches   171;   Indels   106;   Gaps   20:

RESULT 3
KFHU5
coagulation factor V precursor [validated] - human
N; Alternate names: coagulation labile factor; proaccelerin
C; Species: Homo sapiens (man)
C; Date: 19-May-1989 #sequence_revision 02-Jun-1995 #text_change 08-Dec-2000
C; Accession: A56172; A42344; A28028; A27498; A25897
R; Cripe, L.D.; Moore, K.D.; Kane, W.H.
R; Biochemistry 31, 3777-3785, 1992
A; Title: Structure of the gene for human coagulation factor V.
A; Reference number: A56172   MUID:92232668
A; Accession: A56172
A; Molecule type: DNA
A; Residues: 1-2224 <CRI>
A; Cross-references: GB:J05368
A; Accession: A42344
A; Residues: 48-58;79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-5
R; Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Ka
R; Kane, W.H.; Davie, E.W.
R; Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A; Title: Complete cDNA and derived amino acid sequence of human factor V.
A; Reference number: A28028   MUID:87265086
A; Accession: A28028
A; Molecule type: mRNA
A; Residues: 1-857,'R',859-864,'R',866-924,'R',926-1763,'I',1765-2212,'T',2214-2224 <J
A; Cross-references: GB:M16967
A; Accession: A27498
R; Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davie, E.W.
R; Biochemistry 26, 6508-6514, 1987
A; Title: Cloning of cDNAs coding for the heavy chain region and connecting region of
A; Reference number: A27498   MUID:88107560
A; Accession: A27498
A; Molecule type: mRNA
A; Residues: 1-1284,'I',1286-1600 <KAN>
A; Cross-references: GB:M17785
A; Note: parts of this sequence, including the amino end of the mature protein, were d
R; Kane, W.H.; Davie, E.W.
R; Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A; Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor homo
A; Reference number: A25897   MUID:86513665
A; Accession: A25897
A; Molecule type: mRNA
```

A; Residues: 1188-1215,1315-2224 <KA2>
A; Cross-references: GB:M14335
R; Keller, F.G.; Ortel, T.L.; Quinn-Allen, M.A.; Kane, W.H.
A; Note: parts of this sequence were determined by protein sequencing
A; Biochemistry 34, 4118-4124, 1995
A; Reference number: A56139; MUID:95210278
A; Title: Thrombin-catalyzed activation of recombinant human factor V.
A; Contents: annotation; thrombin cleavage sites
A; Comment: Factor V is activated by thrombin and partially by coagulation factor Xa.
A; Genetics:
A; Gene: GDB.F5
A; Cross-references: GDB:119896; OMIM:227400
A; Map position: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1; 6...
A; Introns: 1q23-1q23
A; Function: acts as a cofactor, with calcium and phospholipid, for the factor Xa prot...
A; Description: blood coagulation
A; Pathway: blood coagulation
A; Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re...
A; Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma;
F; 1-28/Domain: signal sequence <SIG>
F; 29-2224/Product: coagulation factor V #status predicted <MAT>
F; 29-737/Product: coagulation factor V heavy chain #status experimental <VAH>
F; 33-329/Domain: A1 <DA1>
F; 346-691/Domain: A2 <DA2>
F; 351-684/Domain: ferroxidase repeat homology <FO1>
F; 692-1573/Domain: B <DDB>
F; 1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F; 1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F; 1574-1905/Domain: A3 <DA3>
F; 1581-1905/Domain: ferroxidase repeat homology <FO3>
F; 1667-1765/Region: phospholipid binding <FO3>
F; 1906-2064/Domain: C1 <DC1>
F; 1906-2061/Domain: discoidin I amino-terminal homology <DN1>
F; 2065-2224/Domain: C2 <DC2>
F; 2065-2221/Domain: discoidin I amino-terminal homology <DN2>
F; 55,239,297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,1479,149...
F; 167-193,248-329,500-526,603-684,1725-1751,1907-2061,2066-2221/Disulfide bonds: #statu...
F; 334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F; 363-693,1546/Binding site: sulfate (Tyr) (covalent) (covalent) #status predicted
F; 376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F; 382-1338/Binding site: carbohydrate (Asn) (covalent) #status absent
F; 534-535/Cleavage site: Arg-Gly (protein C) #status predicted
F; 737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimental
F; 1044-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status experimental
F; 1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match            10.5%;   Score 321;   DB 1;   Length 2224;
Best Local Similarity  42.6%;
Matches  72;  Conservative 30;  Mismatches 51;  Indels 16;  Gaps 5;

```
Qy   300 NGCSRSLSFE-----PDGGIRASS---SWQSVNESGDQVHWSPGQARLQDGPSWASGDSS 352
             ||| ||||||     ||| |||    ||||||||||| || || ||||| ||
Db  2064 NGCSTPLGMENGKIENKQITASSFKKSWW-----GD--TWEPFRARLRNAQGRVRNAWQAKA 2116

Qy   353 NNHKPREMLEIDLGEKKIKTGIRTTGSTQOSNFNFYVKSFVNMFKNNNSKMKTYKGIVNNE 412
             ||:|||||||||:|||||| ||| ::||:|| | |::||||:||| ||| :||:||||
Db  2117 NNNK--QWLEIDLLKIKIKIAIITQOCKSLSSEMVKSYTIHYSEQGVSVNKPYRLKSSMV 2174

Qy   413 EKVFQGNSNFRDPVQVNNFIPPIVARYVRVRVPQTWHQRIALKVELIGCOI 461
             |::|||:||||||||||||||||||:||| |:|::|||||||||||||| |
Db  2175 DKIFEGNTNIKGHVKNFFNPPIISRFIRVIPKTWNQSIALRLELFGCDI 2223
```

RESULT 4

T42764   4
C; Species: Mus musculus (house mouse)
C; Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C; Accession: T42764
R; Yang, T.L.; Cui, J.; Rehumtulla, A.; Yang, A.; Moussalli, M.; Kaufman, R.J.; Ginsburg,...
Blood 91, 4593-4599, 1998

coagulation factor V - mouse

A; Title: The structure and function of murine factor V and its inactivation by prote...
A; Reference number: T42764; MUID:98282202
A; Accession: T42764
A; Molecule type: mRNA; translated from GB/EMBL/DDBJ
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Residues: 1-2183 <TAB>
A; Cross-references: EMBL:U52925; NID:g3219690; PID:g3219691; PIDN:AAC99553.1
A; Function:
A; Pathway: blood coagulation
A; Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
A; Keywords: blood coagulation; glycoprotein; phospholipid binding; plas...
F; 350-602/Domain: ferroxidase repeat homology <FOX1>
F; 1541-1864/Domain: ferroxidase repeat homology <FOX2>

Query Match   10.4%;   Score 318;   DB 2;   Length 2183;
Best Local Similarity 40.8%;   Pred. No. 5.7e-15;
Matches 69; Conservative 31; Mismatches 53; Indels 16; Gaps 5;

```
Qy   300 NGCSRSLSFE-----PDGGIRASS---SWQSVNESGDQVHWSPGQARLQDGPSWASGDSS 352
Db  2023 NGCSTPLGLEDGRIQDKQITASSFKKSWW-----GD--TWEPSLARLRNAQGRVRNAQAKA 2075

Qy   353 NNHKPREMLEIDLGEKKIKTGIRTTGSTQSNFNFYVKSFVNMFKNNSKMKTYKGIVNNE 412
Db  2076 NNNK--QWLQVDLLKIKKVTAIVTQCCKSLSSEMVKSYISIQYSDQGVAWKPYRQKSSMV 2133

Qy   413 EKVFQGNSNFRDPVQVNNFIPPIVARYVRVRVPQTWHQRIALKVELIGCOI 461
Db  2134 DKIFEGNSNTKGHHKNFFNPPIISRFIRIPKTWNQSIALRLELFGCDI 2182
```

RESULT 5

A47285   5
C; Species: Homo sapiens (man)
C; Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C; Accession: A47285
R; Larocca, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Bistrain, A.M.; Ceriani, R.L...
Cancer Res. 51, 4994-4998, 1991
A; Title: A Mr 46,000 human milk fat globule protein that is highly expressed in huma...
A; Reference number: A47285; MUID:91371351
A; Accession: A47285
A; Molecule type: mRNA
A; Status: preliminary
A; Residues: 1-218 <LAR>
A; Cross-references: GB:S56151; NID:g235396; PIDN:AAB19771.1; PID:g235397
A; Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom...
F; 1-56/Domain: discoidin I amino-terminal homology (fragment) <DN1>
F; 60-218/Domain: discoidin I amino-terminal homology <DN2>

milk fat globule protein - human (fragment)

Query Match   9.8%;   Score 299;   DB 2;   Length 218;
Best Local Similarity 37.3%;   Pred. No. 5e-15;
Matches 63; Conservative 33; Mismatches 55; Indels 18; Gaps 5;

```
Qy   300 NGCSRSLSFE-----PDGGIRASSWQSVNESGDQVH----WSPGQARLQDG--PSWASGD 350
Db    59 NGCANPLGLKNNSIPDKQITASSSYKTWG-----LHLFSWNDSYARLDKQGFNFNAMVAGS 113

Qy   351 SSNNHKPREMLEIDLGEKKIKTGIRTTGSTQSNFNTYVKSFVNMFKNNSKMKTYKGIVN 410
Db   114 YGND------QWLQVDLGSSKEVTGITITQGARNFGSVQFVASIVAXSNDSANWTEYQDPRT 169

Qy   411 NEEKVFQGNSNFRDPVQVNNFIPPIVARYVRVQPQTHWAQGRIALKVELGGC 459
Db   170 GSSKIFPGNWMDNHSHKKNLFETPILARVVRILPVAWHNRIALRLELLGC 218
```

RESULT 6

JC4915   6
coagulation factor V precursor - rat
N; Alternate names: O-acetyl-Gd3 ganglioside

C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C:Accession: JC4915
B:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Reference number: JC4915; MUID:96374422
A:Accession: JC4915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
F:1-1;Domain: signal sequence #status predicted <SIG>
F:28-60/Domain: EGF homology <EG1>
F:68-107/Domain: EGF homology <EG2>
F:110-267/Domain: discoidin I amino-terminal homology <DN1>
F:271-427/Domain: discoidin I amino-terminal homology <DN2>
A:Genetics:
A:Gene: ags
C:Genetics:
C:Comment: This protein is required for the O-acetylation of disialoganglioside sialic a
A:Experimental source: CST cell

Query Match   9.7%; Score 297.5; DB 2; Length 427;
Best Local Similarity 26.0%; Pred. No. 1.8e-14;
Matches 91; Conservative 44; Mismatches 96; Indels 119; Gaps 12;

Qy 215 TEYSKFCPA----------GCRDVAGDISGNMVDGYRDTSLLCKNAIHAGIIA-DELGGQI 264
             ||| :| ::: ||| : : |
Db   92 TEYICCPVCPVGYGSGIHCELGCSTKLGLEGGAIAD----SQISASSYVNGNFGLQRNGPEL 146

Qy 265 SVLQRKGI----------SRYEGILANGVLSRDG---------SI, 290
           ||:| :| |: |
Db 147 ARLYRTGIVINAMTASSYDSKPWIQVDFLRKMRVSGVWTQGA-SRAGRAEYLKIFKVAYSL 205

Qy 291 SDKRFLFTS--------- 299
           |||||||
Db 206 DGRRPFTIQDESGTGDKEPWGNQDNNSLKINMNPTLEAQYIRLPYVSCHRGCTLRFELL 265

Qy 300 ----NGCSRSLSFE------PDGQIRASSSWQSVNESGDQVWNSPGQARLQDQGP--SWASG 349
            ||| |||| ||||| |||| || :: :|| | ::
Db 266 GCELHGCSEPLGLKNNNIIPDSGITASSSYKTWNLRA--FGWYPHLGRLDNGGKINAW--- 320

Qy 350 DSSNNHKPREWLEIDLGEKKKIITGIRTTSGTQSNFENFYKSFVVMNFKNNNSKWKTYKGIV 409
            |:: :| |:|: |::|::|||| :::| | :: ::| |: |:|| |
Db 321 -TAQSNSAKEWLQVDLGTQKKVTGIITQGARDFGHIQYVASYKVAHSDDCVQWTVVE--E 377

Qy 410 NNEEKVYFGGNSNFRDDVPQVNFIPPIVARVRVYPQTWHQRIALKVELIGC 459
            |||| || ::||| | |:: | || | :|| || ||
Db 378 QGTSKVFQGNLDNNSHKKNIFEKPFMARVVRVLPLSWHNRITLRLELLGC 427

RESULT 7
A44258
coagulation factor VIII-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: A44258
A:Title: Evidence for a third transcript from the human factor VIII gene.
Genomics 14, 585-589, 1992
R:Levinson, B.; Kenwick, S.; Gamel, P.; Fisher, K.; Gitschler, J.
A:Reference number: A44258; MUID:93052386
A:Accession: A44258
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
C:Cross-references: GB:M90707; NID:g182316; PIDN:AAA58466.1; PID:g182317
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology
F:1-53/Domain: discoidin I amino-terminal homology #status atypical <DN1>
F:57-210/Domain: discoidin I amino-terminal homology; ferroxidase

Query Match   9.4%; Score 285.5; DB 2; Length 216;
Best Local Similarity 37.6%; Pred. No. 5.1e-14;
Matches 62; Conservative 32; Mismatches 58; Indels 13; Gaps 4;

Qy 300 NGCSRSLSFE------PDGQIRASSSWQSVNESGDQVWNSPGQARLQDQGPSWASGDSSSNNH 355
            ||||| ||| |||:||| ||| | ::| |::|| |: : | |||:||
Db 300 NSCSMPLGMESKAISDAGITASSYFTNMFAT------WSRSKARLHLQGRSNAWRPQVNN- 355

Qy 356 KPREWLEIDLGEKKKIITGIRTTSGTQSNFENFYKSFVVMNFKNNNSKWKTYKGIVNNEEKV 415
            ||| |||| |:::||| ::: :| | ::: :| |: |: :|| | :|| | | :|:
Db 110 -PKEWLQVDFQKMMKVTGVTTQGVKSLLTSMYVKEFLISSGSDHSQWQWTLF--FQNGKVKV 166

Qy 416 FGGNSNFRDDVPQVNFIPPIVARVRVYPQTWHQRIALKVELIGCQ 460
            ||| :||| | | | ::| | ||| || | :|
Db 167 FQGNQGSFTPVVNSLDPPLLTRYLRIHPQSMVHQIALRMEVLGCE 211

RESULT 8
E2HU
coagulation factor VIII precursor [validated] - human
C:Alternate names: antihemophilic factor A; coagulation factor VIII:C; procoagulant co
C:Species: Homo sapiens (man)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1995 #text_change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S66445;
C:Cross-references: GB:M88648; NID:g182381; PIDN:AAA52420.1; PID:g182383
R:Wood, W. I.; Capon, D. J.; Simonsen, C. C.; Eaton, D. L.; Gitschler, J.; Keyt, B.; Seeb
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <WOO>
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R:Toole, J. J.; Knopf, J. L.; Wozney, J. M.; Sultzman, L. A.; Buecker, J. L.; Pittman, D.D
s, D. N.; Hewick, R. M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihaemophilic factor.
A:Reference number: I58059; MUID:85061550
A:Accession: I58059
A:Molecule type: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A:Residues: 1-74,'V',76-1259,'E',1261-2351 <RE2>
A:Cross-references: GB:K01740; NID:g182002; PIDN:AAA52484.1; PID:g182803
R:Truett, M. A.; Blacher, R.; Burke, R. L.; Caput, D.; Chu, C.; Dina, D.; Hartog, K.; K
B.; Randolph, A.; Urdea, M. S.; Valenzuela, P.; Dahl, H. H.; Favaloro, J.; Hansen, J.;
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C.
A:Reference number: A23584; MUID:86081164
A:Accession: A23584
A:Molecule type: protein
A:Residues: 1-2351 <TRU>
A:Cross-references: GB:M14113; NID:g182817; PIDN:AAA52485.1; PID:g182818
R:Eaton, D.; Rodriguez, H.; Vehar, G. A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific cleavag
ity.
A:Reference number: A26174; MUID:86159740
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-36;392-399,'X',401-402;1668-1678;1709-1722,'D',1723-1725;1741-1755 <EA
R:Pittman, D. D.; Wang, J. H.; Kaufman, R. J.
Biochemistry 31, 3315-3325, 1992
A:Title: Identification and functional importance of tyrosine sulfate residues within
A:Reference number: A42348; MUID:92207952
A:Accession: A42348

A;Molecule type: protein
R;Residues: 20-36;356-371;392-408;582-594;1668-1669,'X',1671;1672-1692,1693-1708,1709-17
A;Experimental source: recombinant material from NCBI backbone and corrected to correspond with the publi
A;Title: Sequence extracted from Chinese hamster ovary cells
R;Reyte, P.J.; Smudzin, T.M.
J;Biol. Chem. 264, 14005-14010, 1989
R;Gitschier, J.; Wood, W.I.; Goralka, T.M.; Wion, K.L.; Chen, E.Y.; Eaton, D.H.; Vehar,
Nature 312, 326-330, 1984
A;Title: Characterization of the human factor VIII gene.
A;Reference number;; A43986; MUID:89340500
A;Reference number;; A43986; MUID:89340500
A;Contents: annotation
A;Title: Intersubunit fluorescence energy transfer in human factor VIII.
A;Accession: A3986
A;Molecule type: protein
A;Residues: 'X',517-523;1853-1860,'X',1862-1864,'X',1866 <FAY>
R;Leyte, M.; van Schijndel, H.B.; Niehrs, C.; Huttner, W.B.; Verbeet, M.P.; Mertens, K.;
J;Biol. Chem. 266, 740-746, 1991
A;Title: Sulfation of Tyr(1680) of human blood coagulation factor VIII is essential for
A;Reference number;; A56109; MUID:91093266
A;Contents: annotation; sulfation
A;Accession: S63527
A;Molecule type: protein
A;Residues: 733-752;753-759 <KJA>
R;Kjalke, M.; Heding, A.; Talbo, G.; Persson, E.; Thomsen, J.; Ezban, M.
Eur. J. Biochem. 234, 773-779, 1995
A;Title: Amino acid residues 721-729 are required for full factor VIIIa activity.
A;Reference number;; S63527; MUID:96163459
A;Contents: annotation; sulfation
A;Accession: S63527
A;Molecule type: protein
R;Residues: 2,5; spira, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg
R;Lind, P.; Larsson, K.; Spira, J.; Sydow-Baeckman, M.; Almstedt, A.; Gray, E.; Sandberg
Eur. J. Biochem. 232, 19-27, 1995
A;Title: Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction
A;Reference number;; S66445; MUID:96048024
A;Accession: S66445
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1668-1685 <LIN>
A;Comment: Factor VIII is activated by factor Xa and thrombin, but prolonged exposure pr
A;Genetics:
A;Gene: GDB;F8C
A;Cross-references: GDB;119124; OMIM;306700
A;Map position: Xq28-Xq28
A;Introns: 48/2; 89/1; 130/1; 201/1; 224/1; 263/1; 337/1; 424/2; 481/3; 513/1; 584/3; 63
A;Function:
A;Description: acts as a cofactor, with calcium and phospholipid, for the factor IXa pro
A;Pathway: blood coagulation
A;Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
A;Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-2351/Product: coagulation factor VIII #status experimental <MAT>
F;20-740/Product: coagulation factor VIII heavy chain #status experimental <ACH>
F;23-356/Domain: A1 <DA1>
F;192-755/Domain: A2 <DA2>
F;23-348/Domain: A1 <DA1>
F;402-730/Domain: B <DB0>
F;760-1667/Domain: coagulation factor VIIIa light chain #status experimental <ACL>
F;1668-2351/Product: coagulation
F;1709-2038/Domain: A3 <DA3>
F;1716-2038/Domain: ferroxidase repeat homology <FO3>
F;2039-2191/Domain: C1 <DC1>
F;2039-2188/Domain: discoidin I amino-terminal homology <DN1>
F;2192-2351/Domain: C2 <DC2>
F;2192-2345/Domain: discoidin I amino-terminal homology <DN2>
F;60,258,601,776,803,847,919,962,982,1020,1024,1074,1085,1204,1274,1278,1301,1319,1403,1
F;172-198,267-348,547-573,649-730,1851-1877,1918-1922,2040-2188/Disulfide bonds: #status
F;355-356/Cleavage site: Arg-Met (coagulation factor Xa, protein C) #status experimental
F;365,737,738,742,1683,1699/Binding site: sulfate (Tyr) (covalent) #status experimental
F;391-392/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin)

F;414,426/Binding site: sulfate (Tyr) (covalent) #status predicted
F;755-760/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experiment
F;1667-1668/Cleavage site: Arg-Glu (unidentified proteinase) #status experimental
F;1708-1709/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experimi
F;1740-1741/Cleavage site: Arg-Ala (coagulation factor Xa) #status experime
F;2133-2345/Disulfide bonds: #status experimental

Query Match 9.4%; Score 285.5; DB 1; Length 2351;
Best Local Similarity 37.6%; Pred. No. 1.7e-12;
Matches 62; Conservative 32; Mismatches 58; Indels 13; Gaps 4;

Qy 300 NCGCSRSLSFE-------PDCGIRASSWQSVNESGDQVHWSPGQARLQDQGPSWASGCDSSNNH 355
        ||||| |:| ||::|:|:|||:|||::|||||:||||
Db 2191 NSCSMPLGMESKAISDQTQTRASSTFNWFAT------NSPSKARLHLQGRSNARPQVNN- 2244

Qy 356 KPREWLEIDLGERKKIIGIRTTGSTQSNFNFYVKSFYWKPKNNNSKWKTTYKGIVNNEEKV 415
        |||||||||: |||||||||||||||| |:|| |||| |||| |:|||||||||||||||
Db 2245 PKEWLQVDFQKTMKVTGVVTGQVSLISTMSVKEFLISSSSQGHOWTLF--FQNGKVKV 2301

Qy 416 FQGNSNRDPVQONNFPIPIVARYVRVVFQTWHQRIALKVELLGCQ 460
        | || ||:|||:| | ||||||| | ||:| :|::|| ||:|||
Db 2302 FQGNQDSFTPVVNSLDPPLLTRYLRIHHPQSWVHQIALRMEVLGCE 2346

Query Match 9.18%; Score 276.5; DB 1; Length 463;
Best Local Similarity 23.48%; Pred. No. 7.5e-13;
Matches 112; Conservative 59; Mismatches 118; Indels 189; Gaps 21;

Qy 141 CASDYLLFTSSSDQYGPYCGASMTVPKELLLNTSEVTVRFESGSHSGR--------GF 190
        |||| || | :| |: || | : | || :| | | | | | ||
Db 16 CASG--LFAASGD-------FCDS----SLCLN----------GGTCLTGQDNDIYCLCPPGF 55

Qy 191 LITYAS--------SDHFPDLITCLERASHYLATEYSKFCPAGCRDVAGDISGNN- 236
        |: |:|| :|:|:| | ||| |: |:|| | :| |
Db 56 TGLVCNETERGPCSPNPCYNDAKCLVT-LDTQRCDIFTEYICQCPVGYSGCIHCETETNTY 114

Qy 237 -VDG-YRDTSLLCKRAI--------------HAGIIAD------------ 258
        |: |:|| | || | |:| |||
Db 115 NLDGEYMFTTAVENTAVPFTAPTFDLSNNIASRCSTOLGMEGGAIADSOLSASYYVMGFM 174

Qy 259 ---ELGGQISVLQRKGI------------SRYEGILANGVLSRDG--------- 288
        | ||: | |: | | |:| | |:
Db 175 GLGRWGPELARLVRTGIVNAWHASNYDSRKPWIQVNLLRKMRVSGVMTGQA-SRAGRAETL 233

Qy 289 ---SLSDKRFLFTSN---------- 305
        || | :| |:|| | :|:
Db 234 KTFKVAYSLDGRKFEFIQDESGGDKEFLGNLDNNSLKVHMFNPFLEAQYIRLYPVSCHRG 293

Qy 305 -SLSFE---------------------PDCGIRASSWQSVNESGDQVHWSPGQARLQDQG 343

Left column

```
Db    294  CTLRFELLGCELHGCLEPPLGLKNNTIPDSQMSASSSYKTWNLRA--FGWYTLPHLGRLDNQ 351
                : ||        ||   ||   ||  |      || |  || || :      || |  ||
```

RESULT 10

A47004
coagulation factor VIII precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 18-Jun-1999
A:Accession: A47004
A:Molecule type: mRNA
A:Residues: 1-2319 <ELD>
A:Cross-references: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
A:Status: preliminary
A:Reference number: A47004
A:Title: Sequence of the murine factor VIII cDNA.
R:Elder, B.; Lakich, D.; Gitschier, J.
Genomics 16, 374-379, 1993
F:1-19/Domain: signal sequence #status predicted <SIG>
F:23-349/Domain: ferroxidase repeat homology <FOX1>
F:402-730/Domain: ferroxidase repeat homology <FOX2>
F:1686-2005/Domain: ferroxidase repeat homology <FOX3>
F:2007-2156/Domain: discoidin I amino-terminal homology <DN1>
F:2160-2313/Domain: discoidin I amino-terminal homology <DN2>

```
Qy   344  P--SWASGDSSNNHKPREMLEIDLGEKKKITIGRTTGSTQSNFNEYVKSFYWNEFKNNNSK 401
Qy   352  KINAW----TAQSNSAKEMLQVDLGTQRQVVTGIIQGARDFGHIQYVESYKVAHSDDGVQ 407
Qy   402  WKYTKGIVNNEEKYVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELLGC 459
Db   408  WTYVE--EQGSSKVFQGNLDNNSHKKNIFEKPPMARYVRLVPVSWHNRITLRLELLGC 463
```

RESULT 11

Query Match 9.1%; Score 276.5; DB 2; Length 2319;
Best Local Similarity 36.3%;
Matches 61; Conservative 33; Mismatches 61; Indels 13; Gaps 4;

```
Qy   300  NGCGSRSLSFE-----PDGQIRASSSWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSNNH 355
Db  2159  NSCSIPLGMESKVISDTQITASSYFTNMFAT-----WSPSQARLHLQGRTNAWRPQVND-  2212

Qy   356  KPREMLEIDLGEKKKITIGRTTGSTQSNFNEYVKSFYWNNEFKNNNSKWKTYKGIVNNEEKY 415
Db  2213  -PKQWLQVDLQKTMKVTGIITQGVKSLFTSMYVKEFLISSSQDGHHWTQI--LYNGKVKV 2269

Qy   416  FQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELLGCEAQ 463
Db  2270  FQGNQDSSTPPMNSLDPPLLTRYLRIHPQIWEHQIALRLELLGCEAQQ 2317
```

Right column

F:1498-1820/Domain: ferroxidase repeat homology <FOX3>

Query Match 8.9%; Score 271.5; DB 2; Length 2133;
Best Local Similarity 35.9%;
Matches 61; Conservative 29; Mismatches 57; Indels 23; Gaps 5;

```
Qy   300  NGCGSRSLSFE-----PDGQIRASSSWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSNNH 355
Db  1973  NSCSMPLGMQNKAISDSQITRASSHLSNIFAT-----WSPSQARLHLQG-----RINAW 2020

Qy   356  KPR-----EWLEIDLGEKKKITIGIRTTGSTQSNFENFEYVKSFYVMNEFKNNNSKKWKTYKGIVN 410
Db  2021  RPVSSAEBWLQVDLQKTKYVTGITTQGVKSLLSSMYVKEFLYSSSQDGRRWTLF--LQD 2078

Qy   411  NEEKVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELLGCQ 460
Db  2079  GHTKVFQGNQDSSTPVVNALDPFLFTRLRIHPTSWAQHIALRLEVLGCCE 2128
```

RESULT 12

S65138
glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C:Accession: S65138; G48394
R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A:Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal
A:Reference number: S65138; MUID:96125736
A:Accession: S65138
A:Molecule type: mRNA
A:Residues: 1-401 <AOK>
A:Status: preliminary; not compared with conceptual translation
R:Mather, I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-
II-like sequences.
A:Reference number: A48394; MUID:93250576
A:Accession: A48394
A:Molecule type: protein
A:Status: preliminary
A:Residues: 207-220 <MAT>
A:Experimental source: milk
A:Note: sequence extracted from NCBI backbone (NCBIP:131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF hom
C:Keywords: glycoprotein
F:1-32/Domain: EGF homology <EG1>
F:40-79/Domain: EGF homology <EG2>
F:83-239/Domain: discoidin I amino-terminal homology <DN1>
F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 8.8%; Score 268; DB 2; Length 401;
Best Local Similarity 36.8%;
Matches 61; Conservative 30; Mismatches 63; Indels 12; Gaps 4;

```
Qy   300  NGCGSRSLSFE-----PDGQIRASSSWQSVNESGDQVHWSPGQARLQDQGP--SWASGDSSN 353
Db   242  NGCTEPLGLKDNTIPNKQITASSYKTWGLSA--FSWFFYTARLDNQGKRFNAWTAQTNSA 299

Qy   354  NHKPREMLEIDLGEKKKITIGRTTGSTQSNFNEYVKSFYVMNFKNNNSKWKTYKGIVNNEE 413
Db   300  S-----EWLQIDLGSQKRVTGIITQGARDFGHIQVVAAYRVAYGDDGVTWTEYKDPGASES 355

Qy   414  KVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELLGC 459
Db   356  KIFPGNMDNNSHKKNIFETPPQARFVRIQPVAWHINRITLRVELGG 401
```

RESULT 13

T42763
coagulation factor VIII precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000
C:Accession: T42763
R:Lollar, P.
submitted to the EMBL Data Library, August 1996
A:Reference number: T42763
A:Accession: T42763
A:Molecule type: mRNA
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Residues: 1-2133 <LOL>
A:Cross-references: EMBL:U49517; NID:g1511633; PID:g1511634; PIDN:AAB06705.1
C:Superfamily: coagulation factor VIII; discoidin I amino-terminal homology; ferroxidase
C:Keywords: acute phase; blood coagulation; duplication; glycoprotein; hemophilia A; pla
A:Molecule type: mRNA
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-2133/Product: coagulation factor VIII #status predicted <MAT>
F:23-349/Domain: ferroxidase repeat homology <FOX1>
F:402-730/Domain: ferroxidase repeat homology <FOX3>

Query Match 8.9%; Score 271.5; DB 2; Length 2133;
Best Local Similarity 35.9%; Mismatches ;

RESULT
S74211

```
F:6-40/Domain: EGF homology <EGF>

PAS-6/7 protein precursor - bovine
C;Species: Bos prinigenius taurus (cattle)
C;Date: 04-Dec-1997
R;Hvarregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
A;Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk glob
A;Reference number: S74211; MUID:97008954
A;Accession: S74211 <HVA>
A;Molecule type: mRNA
A;Residues: 1-427 <HVA>
A;Cross-references: EMBL:X91895; NID:g1632778; PIDN:CA62997.1; PID:g1632779
A;Accession: S78114
A;Molecule type: protein
A;Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427
R;Kim, D.H.; Kanno, C.; Mizokami, Y.
A;Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
Biochim. Biophys. Acta 1122, 203-211, 1992
Eur. J. Biochem. 240, 628-636, 1996
A;Reference number: S23926; MUID:92355107
A;Accession: S24181
A;Molecule type: protein
A;Residues: 383-394 <XIM>
F;6-105/Domain: EGF homology <EG2>
F;108-265/Domain: discoidin I amino-terminal homology <DN1>
F;269-427/Domain: discoidin I amino-terminal homology <DN2>
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-427/Product: PAS-6/7 protein #status experimental <MAT>
F;24-35,29-47,49-58,66-77;71-94,96-105/Disulfide bonds: #status experimental
F;27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;59,227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;109-265,252-256,270-427/Disulfide bond: glycoprotein; milk
Keywords: blocked amino end; disulfide bond; glycoprotein; EGF homolo
Superfamily: milk fat globule protein, discoidin I amino-terminal homology; EGF homol

Query Match       8.8%;      Score 268;      DB 2;      Length 427;
Best Local Similarity    36.7%;   Pred. No. 2.9e-12;
Matches   61;   Conservative   30;   Mismatches   63;   Indels   12;   Gaps   4;

Qy   300  NGCSRSLSFE-----PDCGIRASSWQSVVNESGDQVHWSPGQARLQDQGP--SWASGDSSN 353
          |||:           ::||:|: ||||: :: :|| |::|| ||:||   |||||||
Db   268  NGCTEPLGLKDNTIPPNKQITASSYKTWGLSA--FSWPYVARLDNQGKFNAWTAQTNSA 325

Qy   354  NHKPREWLEIDILGEKKKITGIRTGSTQGSNFNFYVKSFVNNFKNNNSKWKKTYKGIVNNEE 413
          ||:|||||||||:|||||||||:|:|||||||| ||  :: :  |: : : | |:
Db   326  S-----EWLQIDLGSQKRVTGIITQGARDFGHIQYVAAYRVAYKDDGVTWTEYKDPCGASES 381

Qy   414  RVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELIGC 459
          |:|| || |||:||||||||:|||||:| |||:|||:|||||:||
Db   382  KIFPGNMDNSHKKNIFETPFQ0ARFVRIQPVAWHNRITLRVELLGC 427
```

```
RESULT 14
T11743
pP47 protein - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
A;Accession: T11743
R;Ensslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
Biol. Reprod. 58, 1057-1064, 1998
A;Title: Molecular cloning and characterization of P47, a novel boar sperm-associated zo
A;Reference number: Z17325; MUID:98206817
A;Accession: T11743
A;Molecule type: mRNA
A;Residues: 1-409 <ENS>
A;Cross-references: EMBL:Y11683; NID:g2652927; PIDN:CAA72379.1; PID:g2652928

Query Match       8.5%;      Score 259.5;      DB 2;      Length 409;
Best Local Similarity    25.0%;   Pred. No. 1.2e-11;
Matches   87;   Conservative   46;   Mismatches   102;   Indels   113;   Gaps   11;

Qy   215  TEYSKFCPACGRVAGADISGNMVDGYRDTSL-----LCKAAHHAGII----ADELGG--- 262
          |:||:| || || |:|:|||||:|  :| ::     ||:|||| ||     ||::|
Db   72   TEYICKCPHGYGTIHCEIICNAPLGMEGTGA IADPICASSWHI.GFWGIQRWAPELARLHR_13¶

Qy   263  ------------QISVLQRGISRYEGILANCYLSRDCSL- 290
          :::|  | |:: :| :::|||
Db   132  AGIVANWTAASNYDRNPWIQVNLLRR--MRVTCVVTGQA-SRAGSAEYMKTFKVAYSTDG 187

Qy   291  -----SDKRFL------------ 302
          :|||:|
Db   188  RKFQFIQGAEESGDKIFMGNLDNSGLKVNLFEVPLEVQVRLVLPIICHRGCTLRFELLGC 247

Qy   303  SRSLSFEPDG------------QIRASSWQSVVNESGDQVHWSPGQARLQDQGP--SWASGDS 351
          ||:|| ||             ::||:|: ||||: :: :|| |::||
Db   248  ELSGCAEPLGLKIDNTIPNKQITASSYFIWGLSA--FSWPYTARLDNQGKFNAWTAQTAQSN 305

Qy   352  SNNHKPREWLEIDILGEKKKITGIRTGSTQGSNFNFYVKSFVNNFKNNNSKWKKTYKGIVN 411
          |||:||||||||:|||||||||:|:|||||||| ||  :: :  |: : : | |:
Db   306  SAS----EWLQIDLGSQRRVTGIITQGARDFGHIQYVAAYKVAYSDGVSWTEYRDQGAL 361

Qy   412  EERVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELIGC 459
          |:|| || |||:||||||||:|||||:| |||:|||:|||||:||
Db   362  EGKIFPGNLDNNSHKKNMFETPFLTRFVRILPVAWHNRITLRVELLGC 409
```

```
RESULT 15
JC5256
adipocyte transcription factor, AEBP1 - human
C;Species: Homo sapiens (man)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
R;Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo,
Biochem. Biophys. Res. Commun. 228, 411-414, 1996
A;Title: A cDNA cloning of human AEBP1 from primary cultured osteoblasts and its expr
A;Reference number: JC5256; MUID:97079196
A;Accession: JC5256
A;Molecule type: mRNA
A;Residues: 1-845 <OHN>
A;Cross-references: DDBJ:D86479; NID:g1468942; PIDN:BAA13094.1; PID:g1468943

Query Match       7.1%;      Score 217;      DB 2;      Length 845;
Best Local Similarity    25.7%;   Pred. No. 5.3e-08;
Matches   97;   Conservative   40;   Mismatches   97;   Indels   22;   Gaps   5;

Qy   311  DCGIRASSWQ-SVNESGDQVHWSPGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKK 369
          |||:| || :  :: || |::||||| |:||  |||||    :   ::| |||:|||
Db   84   DNGIRASSMLRHGLGAQRGRLNMQTGATEDDYDGACAEDDART----QWIEVDTRRT 139

Qy   370  KITGIRTGSTQGSNFNFYVKSFVNNFKNNNSKWKKTYKGIVNNEEKVFQGNSNFRDPVQNN 429
          |||||:|:|||||||| ||  :: :  |: : : | |:
Db   140  RFTGVITQGRDSSIHODFVTFFVGFGNSDSOTWVWYTN--GYEEMTFHGNVKDFDPVLSE 197

Qy   430  FIPPIVARYVRVVPQTWHQRIALKVELIGCQQITQ----------GNGSLVWR----KTS 474
          ||||:|||||:| |||:|||:|||||:||:||:|             ||::|     | |
Db   198  LPEPVVARFIRIYPLWNGSLCMRLEVLGCSVAPVYSYTAQNEVVATDLDFRHHSYKDM 257

Qy   475  QSTSVSTKKDEITRPIPSEEFTSTGINITTVAI 508
          :| : : : :    :: :::   : | ||
Db   258  RQLMKVVNEECPFITRTYSLGKSSRGLKIYAMEI 291
```

Search completed: July 9, 2002, 02:16:44
Job time: 3801 sec

```
OM protein - protein search, using sw model

                    GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:             July  9, 2002, 02:15:23 ; Search time 35.54 Seconds
                                              (without alignments)
                                              638.426 Million cell updates/sec

Title:              US-09-691-344A-4
Perfect score:      3052
Sequence:           1 MGFGAQGRLRFVPAPRSSAE.........YDNEKEMTQKLDLITSDMAG 586

Scoring table:      BLOSUM62
                    Gapop 10.0 , Gapext 0.5

Searched:           105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters:     105224

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :          SwissProt_40:*


Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.


                              SUMMARIES

Result          Query
No      Score   Match   Length  DB      ID              Description
------  -----   -----   ------  --      --------------  -----------
  1     351.5   11.5    925     1       NRP2_RAT        O35276; rattus norv
  2     349     11.5    931     1       NRP2_HUMAN      O60462; homo sapien
  3     343.5   11.3    931     1       NRP2_MOUSE      O35375; mus musculu
  4     338     11.1    914     1       NRP1_CHICK      P79796; gallus gall
  5     335     11.0    923     1       NRP1_HUMAN      O14786; homo sapien
  6     335     10.8    922     1       NRP1_RAT        Q9QWJ9; rattus norv
  7     331     10.8    2211    1       NRP1_BOVIN      Q28107; bos taurus
  8     331     10.8    923     1       NRP1_XENLA      P28824; xenopus lae
  9     330     10.8    923     1       NRP1_MOUSE      P97333; mus musculu
 10     329     10.8    2224    1       FA5_HUMAN       P12259; homo sapien
 11     317     10.3    2258    1       FA5_PIG         Q9I9P1; sus scrofa
 12     314     10.4    387     1       MFGM_HUMAN      Q08431; homo sapien
 13     299     9.7     427     1       MFGM_RAT        P70490; rattus norv
 14     297.5   9.7     427     1       FA8_HUMAN       P00451; homo sapien
 15     285.5   9.4     2351    1       MFGM_MOUSE      P21956; mus musculu
 16     276.5   9.1     463     1       MFGM_MOUSE      Q6194; mus musculu
 17     271.5   8.9     2319    1       FA8_MOUSE       P12263; mus musculu
 18     268     8.8     2133    1       FA8_PIG         P12264; sus scrofa
 19     259.5   8.8     409     1       MFGM_BOVIN      Q95114; bos taurus
 20     259.5   8.5     427     1       MFGM_PIG        Q7938S; sus scrofa
 21     225     7.3     280     1       XLR1_FUGRU      Q9w6r5; fugu rubrip
 22     221.5   7.3     224     1       XLR1_MOUSE      Q9zl14; mus musculu
 23     218.5   7.2     224     1       XLR1_HUMAN      O15557; homo sapien
 24     197     6.5     1308    1       CTA4_HUMAN      P00431; homo sapien
 25     195.5   6.4     1331    1       CTA2_HUMAN      P00754; homo sapien
 26     194.5   6.4     1310    1       CTA4_MOUSE      Q99p47; mus musculu
 27     192.5   6.3     1284    1       CTA2_MOUSE      Q94807; drosophila
 28     187     6.1     3133    1       HMCT_BOMMO      Q94887; bombyx mori
 29     186.5   6.1     3133    1       NRX4_DROME      P98092; bombyx mori
 30     184.5   6.0     686     1       MAS2_HUMAN      O00187; homo sapien
 31     172.5   6.0     1154    1       CTA3_HUMAN      Q9b276; homo sapien
 32     165     5.4     514     1       UVS2_XENLA      P42564; xenopus lae
 33     165     5.3     986     1       BMP1_HUMAN      P13497; homo sapien
 32     163     5.3     986     1       PCO1_RAT        O08628; rattus norv
 33     162.5   5.3     704     1       CRAR_MOUSE      P98064; mus musculu
```

```
RESULT    1

ID      NRP2_RAT        STANDARD;       PRT;     925 AA.
AC      O35276;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE      receptor 2).
GN      NRP2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SPRAGUE-DAWLEY;
RX      MEDLINE=97433085; PubMed=9288754;
RA      Kolodkin A.L., Levengood D.V., Rowe E.G., Tai.Y.-T., Giger R.J.,
RA      Ginty D.D.;
RL      "Neuropilin is a semaphorin III receptor.";
RL      Cell 90:753-762(1997).
CC  -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC      AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC  -!- TISSUE SPECIFICITY: FOUND IN CERTAIN NEURONAL POPULATIONS OF THE
CC      CNS AND IN OTHER NONNEURONAL TISSUES INCLUDING MESENCHYMAL TISSUE
CC      LINING IN THE RIBS.
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC  -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC  -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC  -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC  --------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  --------------------------------------------------------------------------
DR      EMBL; AF016297; AAC53338.1; -.
DR      InterPro; IPR000859; CUB.
DR      InterPro; IPR000421; FA58_C.
DR      InterPro; IPR000998; MAM.
DR      Pfam; PF00431; CUB; 2.
DR      Pfam; PF00754; F5_F8_type_C; 2.
DR      Pfam; PF00629; MAM; 1.
DR      SMART; SM00042; CUB; 2.
DR      SMART; SM00231; FA58C; 2.
DR      SMART; SM00137; MAM; 1.
DR      PROSITE; PS01180; CUB; 2.
DR      PROSITE; PS01285; FA58C_1; 2.
DR      PROSITE; PS01286; FA58C_2; 2.
DR      PROSITE; PS00060; MAM_1; 1.
DR      PROSITE; PS50060; MAM_2; 1.
KW      Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.
```

ALIGNMENTS

```
34      162.5   5.3     854     1       DDR2_MOUSE      Q62371 mus musculu
35      161.5   5.3     449     1       PCO1_HUMAN      Q15113 homo sapien
36      160     5.2     468     1       PCO1_MOUSE      Q61398 mus musculu
37      159     5.2     550     1       COCH_HUMAN      O43405 homo sapien
38      158     5.2     991     1       BMP1_MOUSE      P98063 mus musculu
39      158     5.1     1022    1       TLD_BRARE       O57460 brachydanio
40      157     5.1     552     1       COCH_MOUSE      Q62507 mus musculu
41      156     5.1     707     1       BMP1_XENLA      P99070 xenopus lae
42      156     5.1     855     1       DDR2_HUMAN      Q16832 homo sapien
43      155.5   5.1     695     1       CASP_MESAU      Q16833 mesocricetu
44      146.5   4.9     1019    1       LFC_TACTR       P28175 tachypleus
45      142.5   4.7     1385    1       CTAI_MOUSE      O54991 mus musculu
```

```
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    925       NEUROPILIN-2.
FT   DOMAIN       23    858       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    859    883       POTENTIAL.
FT   DOMAIN      884    925       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       28    142       CUB 1.
FT   DOMAIN      149    267       CUB 2.
FT   DOMAIN      277    427       F5/8 TYPE C 1.
FT   DOMAIN      434    592       F5/8 TYPE C 2.
FT   DOMAIN      642    802       MAM.
FT   DISULFID     28     55       BY SIMILARITY.
FT   DISULFID     83    105       BY SIMILARITY.
FT   DISULFID    149    175       BY SIMILARITY.
FT   DISULFID    208    230       BY SIMILARITY.
FT   DISULFID    277    427       BY SIMILARITY.
FT   DISULFID    434    592       BY SIMILARITY.
FT   CARBOHYD    152    152       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    157    157       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    833    833       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    834    834       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    925 AA;  103896 MW;  3BF62903F644851C CRC64;
```

```
Qy  Query Match    11.5%;   Conservative   24.4%;   Pred. No. 7.2e-19;
    Best Local Similarity                Score 351.5;  DB 1;   Length 925;
    Matches 114;   Conservative 76;   Mismatches 181;   Indels 97;   Gaps 17;

Qy    88  CGHLVTYQDSGTMTSKNYPGTYPNITVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY  145
          | :||  | |  | || :|| ||||   || |     |::|:|:   | ::|  || |
Db    28  CGGRLNSKDAGYITSPGFPDQYPSHQNCENVVYAPENQKIVLNFNPHFEIEKHDCKYDF   87

Qy   146  LLF------TSSDQYGYPYCGSMYVPKELLLNTSERVTVRFESGSHISGRGFLLTY-----AS  196
          ||:       :|  :|  | | :  ||| :|| | :   :| || |:| :| ::|       ||
Db    88  IEIRDGDSDSADLLGKHCGNI-APPTIISSGSVLYIKFTSDYARQGAGFSLRYEIFKTGS  146

Qy   197  SD------------HPDLITC-----LERASHYLKTEYSKY---------------QS  220
          :|             ||| |       | | :  |   |                |
Db   147  EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPRMEIILQFLTFDLEHDPLQVGEG  206

Qy   221  -CPAGCRDVACDIS--GNNVDGYRDTSLLCKAAIHAGIIADELGQQISVLQRKGISRY--  275
           | | |||:  |:    ||:|||| ||  |      |:|:    |  ||::|: || |
Db   207  DCKYDMLDIWDGIPHVGPLGIKYCGTKTPSKLRSSTGILSIFHYDMAVAKDCGFSARYYL  266

Qy   276  ------EGILANGVLSRD-GSLSDKREFLFTSNGCSRSLSFEPDQGIRASSWQSVSNESGD  328
                  | : | |  ||  |:  :  | :  |                | :||:  :   |
Db   267  VHQEPENFQCNAPLGMESGRIANE------------------------QISASSTF-----  300

Qy   329  QVHNSPGQARLQDQGPSVNASGDSSNNHKPREWLEIDLGEKKKIITGIIRTG--STQSNFNF  386
          |   | | :||  :  :||     ::    ||   :|| : :  | |   |    |::  :
Db   301  DGRWTPQQSRLHGDDNGWTPNVDSN----KEYLQVDLRFLTMLTAIATQGAISRETQKGY  356

Qy   387  YVKSFVMNFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARVVRVVPQTW  446
          ||||::: :  ::  |  | :    :| ||| |:    :|   :: |: ||::|:||||:|
Db   357  YVKSYLEVSTNGEDMWVYRHGKNH--KVFQANNDATELVLNKLHTPLLTRFIRIRPQTW  414

Qy   447  HQRIALKVELIGCQITQGNDSLVWRKYTS---QSTLSVSTVKKEDETTTRP  491
          | :||| :||  | :|     : |:       ||::|:|:      :  |
Db   415  HLGIALRLELEFGCRVTDAPCSNMLGMLSGLIADTQISASSTREYLMSP  462
```

```
RESULT 2

ID   NRP2_HUMAN     STANDARD;      PRT;   931 AA.
AC   O60462; O14820; O14821;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE   receptor 2).
GN   NRP2 OR VEGF165R2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A0 AND A17).
RX   MEDLINE=97470888; PubMed=9311348;
RA   Chen H., Chedotal A., He Z., Goodman C.S., Tessier-Lavigne M.;
RT   "Neuropilin-2, a novel member of the neuropilin family, is a high
RT   affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT   III.";
RL   Neuron 19:547-559(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM A22).
RX   MEDLINE=98188099; PubMed=9529250;
RA   Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT   "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT   isoform-specific receptor for vascular endothelial growth factor.";
RL   Cell 92:735-745(1998).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE=20309748; PubMed=10748121;
RA   Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT   "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT   form of vascular endothelial growth factor (VEGF) and of placenta
RT   growth factor-2, but only neuropilin-2 functions as a receptor for
RT   the 145-amino acid form of VEGF.";
RL   J. Biol. Chem. 275:18040-18045(2000).
CC   -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC       AND VEGF-145 ISOFORMS OF VEGF, AND THE PLGF-2 ISOFORM OF PGF.
CC   -!- SUBUNIT: NEUROPILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC       NEUROPILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN 3C RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A22 (SHOWN HERE), A0 AND A17;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF022859; AAC51788.1; -.
DR   EMBL; AF022860; AAC51789.1; -.
DR   EMBL; AF016098; AAC12922.1; -.
DR   MIM; 602070; -.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR000998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   Pfam; PF00629; MAM; 1.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
DR   PROSITE; PS50060; MAM; 1.
KW   Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW   Alternative splicing.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    931       NEUROPILIN-2.
FT   DOMAIN       21    864       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    865    889       POTENTIAL.
FT   DOMAIN      890    931       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       28    142       CUB 1.
FT   DOMAIN      149    267       CUB 2.
```

```
FT   DOMAIN       277    427   F5/8 TYPE C 1.
FT   DOMAIN       434    592   F5/8 TYPE C 2.
FT   DOMAIN       642    802   MAM.
FT   DOMAIN       671    674   POLY-SER.
FT   DISULFID      28     55   BY SIMILARITY.
FT   DISULFID      83    105   BY SIMILARITY.
FT   DISULFID     149    175   BY SIMILARITY.
FT   DISULFID     208    230   BY SIMILARITY.
FT   DISULFID     277    427   BY SIMILARITY.
FT   DISULFID     434    592   BY SIMILARITY.
FT   CARBOHYD     152    152   N-LINKED (GLCNAC...).
FT   CARBOHYD     157    157   N-LINKED (GLCNAC...).
FT   CARBOHYD     629    629   N-LINKED (GLCNAC...).
FT   CARBOHYD     839    839   N-LINKED (GLCNAC...).
FT   CARBOHYD     813    813   N-LINKED (GLCNAC...).
FT   VARSPLIC     809    809   MISSING (IN ISOFORM A1?).
FT   VARSPLIC     809    830   MISSING (IN ISOFORM A0).
FT   CONFLICT     602    602   E -> K (IN REF. 1).
FT   SEQUENCE     931 AA;  104830 MW;  270CBAE69A0A797C CRC64;
```

```
Query Match          11.4%;
Best Local Similarity  25.3%;                    Score 349;   DB 1;   Length 931;
Matches  109;  Conservative  72;   Pred. No. 1.1e-18;
Mismatches  166;   Indels  84;   Gaps  16;
```

```
Qy   88  CCHLVTYQDSGCTMTSKNYPCGTYPFHTVCEKTITVPR-GKRLLRLG-DLDIESQTCASDY  145
         :|||:|||::|:   :|||:   :|||:|  : :: ::|  |:|:|:| :||::  : :|
Db   28  CGGRLNSKDAGTITSPGYPODYFPSHONCEWIVYAPEPNQKIVLNFNPHFEIEKHDCKYDE   87

Qy  146  LLF----TSSSDQYGPYCGSMTVPKEILLNTSEVTVRPESSGSHISGRGFLITY----AS  196
         |  |    ::||| :   ::|| :: ||||:|| :: : : :|:|::: :| ::|    |:
Db   88  IEIRDGDSESADLLGKHCGNI-APPTIISSGSMLYIKFTSDYARPGAGFSLRYEIFKTGS  146

Qy  197  SD---------------HPDLITC----LERASHYLKTEYSKF-----------------  220
         |:               ||: :|     ||:|:| :| ::|
Db  147  EDCSKSNNTSPNGTIESPGFPEKYPHNLDCTFTILANRKMEIIIQFLIFPLEHDPLQVGEG  206

Qy  221  -CPAGCNRDVAGDIS--GNMVDGYRDPSLLCKAAIHAGIIADELGGGISVLORKGISRYEG  277
          :||| |: ::|    ||  |: | :|: ||:|||| :: :|||||:: ||:|:|:||
Db  207  DCXYDWLDIWDGIPHVGPLIGKYCGTKYPSELRSSTGILSLTFHTDMAVAK-------DG  259

Qy  278  ILANGVLSRDGSLSDKRFLETSNGCSRSLSFE----PDGQIRASSSWQSVNESGDQVHVS  333
         |:||  | :|    ::| |:|||    |:|      |||:|| :: :|:|| :|||::|
Db  260  FSARYLVLHQEPLENFQ-------CNVPLGMESGRIANEQISASSTY-----SDGRWT  305

Qy  334  PQOARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKTIGIRTTG---STQSNFNFYVKSF  391
         |:| ||  :| :|     ::|    : |:|:||   : |    :|    |  |:| :|||:
Db  306  POQSRLGDDGNGWTPNLDSN----KEYLQVDLRFLTMLTATATQGAISRETGNGIYVKSY  361

Qy  392  VNNFKNNNSKWKTYKGIVNNBEKVFQGNSNFRDPVQONNFIPPIVARYVRVPQTWHGRIA  451
         :: |:|      |  |   :  :|||:::  |::| :|: |   ::|:|||:||||:|||
Db  362  KLEVSTNGEDDWMVYRHGKNH---KVFQANNDAIYEVVLNKLHAPLLTRFVRIRPQTWHSGIA  419

Qy  452  LKVELIGCQIT  462
         |::|| ||::|
Db  420  LRLELFGCRVT  430
```

RESULT 3

```
ID   NRP2_MOUSE     STANDARD;      PRT;   931 AA.
AC   O35375; O35373; O35374; O35376; O35377; O35378;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neuropilin-2 precursor (Vascular endothelial cell growth factor 165
DE   receptor 2).
GN   NRP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
```

```
RC   STRAIN=BALB/C;
RX   MEDLINE=97470888; PubMed=9331348;
RA   Chen H., Chedotal A., He Z.-G., Goodman C.S., Tessier-Lavigne M.;
RT   "Neuropilin-2, a novel member of the neuropilin family, is a high
RT   affinity receptor for the semaphorins Sema E and Sema IV but not Sema
RT   III.";
RL   Neuron 19:547-559(1997).
CC   -!- FUNCTION: HIGH AFFINITY RECEPTOR FOR SEMAPHORINS 3C, 3F, VEGF-165
CC       AND VEGF-145 ISOFORMS OF VEGF AND THE PLGF-2 ISOFORM OF PGF
CC       (POTENTIAL). NEUROPILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC   -!- SUBUNIT: NEUROPILIN-2 PROBABLY FORMS AN HETEROMERIC COMPLEX WITH
CC       NEUROPILIN-1 IN ORDER TO BE A FUNCTIONAL SEMAPHORIN RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 6 ISOFORMS; A22 (SHOWN HERE), AO, A5, A17,
CC       A0 AND B5; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING CNS, PNS AND IN SOME
CC       NONNEURAL TISSUES INCLUDING LIMB BUDS, DEVELOPING BONES, MUSCLES,
CC       INTESTINAL EPITHELIUM, KIDNEY, LUNG AND SUBMANDIBULAR GLAND.
CC   -!- DEVELOPMENTAL STAGE: THE EXPRESSION PATTERN IS VERY DYNAMIC AND
CC       IS DEVELOPMENTALLY REGULATED.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AF022856; AAC53379.1; -.
DR   EMBL; AF022854; AAC53377.1; -.
DR   EMBL; AF022855; AAC53378.1; -.
DR   EMBL; AF022857; AAC53380.1; -.
DR   EMBL; AF022858; AAC53381.1; -.
DR   EMBL; AF022861; AAC53382.1; -.
DR   MGD; MGI:1100492; Nrp2.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR000998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   Pfam; PF00629; MAM; 1.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C; 2.
DR   PROSITE; PS01286; FA58C; 2.
DR   PROSITE; PS00060; MAM; 1.
KW   Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW   Alternative splicing.
FT   SIGNAL         1     20
FT   CHAIN         21    931
FT   DOMAIN        21    864   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     865    889   POTENTIAL.
FT   DOMAIN       890    931   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        28    142   CUB 1.
FT   DOMAIN       149    267   CUB 2.
FT   DOMAIN       277    427   F5/8 TYPE C 1.
FT   DOMAIN       434    592   F5/8 TYPE C 2.
FT   DOMAIN       642    802   MAM.
FT   DOMAIN       671    845   POLY-SER.
FT   DISULFID      28     55   BY SIMILARITY.
FT   DISULFID      83    105   BY SIMILARITY.
FT   DISULFID     149    175   BY SIMILARITY.
FT   DISULFID     208    230   BY SIMILARITY.
FT   DISULFID     277    427   BY SIMILARITY.
FT   DISULFID     434    592   BY SIMILARITY.
FT   CARBOHYD     152    152   N-LINKED (GLCNAC...).
FT   CARBOHYD     157    157   N-LINKED (GLCNAC...).
```

```
FT   CARBOHYD    629    629       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    839    839       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    809    813       MISSING (IN ISOFORM A17).
FT   VARSPLIC    809    830       MISSING (IN ISOFORM A0).
FT   VARSPLIC    814    830       MISSING (IN ISOFORM A5).
FT   VARSPLIC    810    931       MISSING (IN ISOFORM B3).
```

```
FT   CONFLICT    786    786       G -> S (IN REF. 1; AAC53380 AND
FT                                AAC53381).
SQ   SEQUENCE   931 AA;  104558 MW;  76F2443F411D2F63 CRC64;
```

```
Qy         931  SEQUENCE   931 AA;  104558 MW;  72; Mismatches 177; Indels 119; Gaps 15;
```

```
Query Match                11.3%;  Score 343.5;  DB 1;  Length 931;
Best Local Similarity      23.2%;  Pred. No. 3e-18;
Matches 111; Conservative   72; Mismatches 177; Indels 119; Gaps 15;

Qy    88   CGHLVTYQDSGTMYRSKNYFPGTYPNHTVCEKTITYVPK-GKRLILRLG-DLDIESGQTCASDY   145
Db    28   CGGRPNSKDAGYIISPGYPQDYPSHQNCEWIVYAPEPNQKIVLNFNPHFFIEKHDCKYDF    87
Qy    146  LLF-----TSSSDQYGFYCGSMTVPKRELLINTSEVTVRFEGSGSHISGRGFLTY-----AS   196
Db    88   IEIRDGDSESADILGKHGNI-APPTIISSGSVLIKFTSDYARQGAGFSLRYEIFKTGS   146
Qy    197  SD----------------------HPDLITC---LERASHYLKTEYSKF-------   220
Db    147  EDCSKNFTSPNGTIESPGPEKYPHNLDCTFTILAKPRMEIILQFLTEDLEHDPLQVGEG   206
Qy    221  -CPAGCRDVAGDIS--GNWVDGYRDTSLJCKAAIHAGIIADELGGGISVLORKGISRY--   275
Db    207  DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSKLRSSTGILSLTEHTDMAVAKDGFSARYYL   266
Qy    276  ----EGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGQIRAS   317
Db    267  IHQEPENFPQCNVPLGMESGRIANEQISASSTFSDGR----------   303
Qy    318  SSWQSVNESGDQVHWSPGQARLQDGGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTT'   377
Db    304  ------------WTPQQSRLHGDDNGWTPNLDSN---KEILQVDLRFLTMLTMLTAIATQ   345
Qy    378  G--STGSNFNFYVKSFVMNFKNNNSKWKTVKGIVNNEEKVFQCGNSNFRDPVQONNFIPPIV   435
Db    346  GAISRETQKGIYVASKYLEVSTNGEDWMVYRHGKNH--KIFQANNDAVEFVLNKLHHPLL   403
Qy    436  ARVRVRVPQTWHQRIALKVELIGCOITGGNDSLVWRKTS---QSTSVSTKEDETITRP   491
Db    404  TRFIRIRPQTWHLGIALRLELFGCRVDAPCSNMLGMLSGLIADTGISASSTREYLMSP   462
```

```
RESULT 4
ID   NRP1_CHICK   STANDARD;   PRT;   914 AA.
AC   P79795;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neuropilin-1 precursor (A5 protein).
GN   NRP1 OR NRP.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=Embryonic brain;
RX   MEDLINE=95324761; PubMed=7601310;
RA   Takagi S., Kasuya Y., Shimizu M., Matsuura T., Tsubol M., Kawakami A.,
RA   Fujisawa H.;
RL   "Expression of a cell adhesion molecule, neuropilin, in the
RL   developing chick nervous system.";
RL   Dev. Biol. 170:207-222(1995).
```

```
CC   -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC       CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC       CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
CC       SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
CC       SIMILARITY). SEEMS TO HAVE CALCIUM-INDEPENDENT CELL ADHESION
CC       PROPERTIES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: DEVELOPING NERVOUS SYSTEM; OPTIC TECTUM
CC       (LAYERS D AND E OF SGFS), AMACRINE CELLS OF RETINA, NEURITES OF
CC       DORSAL ROOT GANGLIA. ALSO EXPRESSED IN NONNEURONAL CELLS, E.G.
CC       BLOOD VESSELS IN THE ENTIRE EMBRYO.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; D45416; BAA08256.1; -.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR000998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   Pfam; PF00629; MAM; 1.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS50060; MAM_2; 1.
```

```
KW   Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW   Cell adhesion.
FT   SIGNAL        1     18
FT   CHAIN        19    914       POTENTIAL.
FT   DOMAIN       20    145       NEUROPILIN-1.
FT   TRANSMEM    848    870       POTENTIAL.
FT   DOMAIN      871    914       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     25    139       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      145    263       CUB 1.
FT   DOMAIN      273    422       CUB 2.
FT   DOMAIN      429    581       F5/8 TYPE C 1.
FT   DOMAIN      429    581       F5/8 TYPE C 2.
FT   DOMAIN      636    801       MAM.
FT   DISULFID     25     52       PROBABLE.
FT   DISULFID     80    102       PROBABLE.
FT   DISULFID    145    171       PROBABLE.
FT   DISULFID    204    226       PROBABLE.
FT   DISULFID    273    422       BY SIMILARITY.
FT   DISULFID    429    581       BY SIMILARITY.
SQ   SEQUENCE   914 AA;  102480 MW;  DD2EE6D6F0CBB68C CRC64;
```

```
Query Match               11.1%;  Score 338;  DB 1;  Length 914;
Best Local Similarity     25.7%;  Pred. No. 7.6e-18;
Matches 115; Conservative  67; Mismatches 180; Indels 86; Gaps 17;
```

RESULT 5

```
Qy   68  ALLLAVSAPLRLQAEELGDGCGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPKG-KR  126
             : |||  :    ::  ::                                 :  |::| |:
Db   11  ALTFTLSRALR------SDKCGDPTIKILSPGYLITSPGYLPQSYHPSQKCEWLLQAPEPVQR  64

Qy  127  LILRLG-DLDIESQTCASDYILFTSSSDQ---YGPYCGSMTVPKELLLNTSEVTVRFES  181
          ::: :  :|  ::   |  :: :: : ::   :|  : : :  ::::       :: |
Db   65  IMFNFNHFDLEDRDKYDVEVIDGDNAEGRLWGKYKYCGKI-APPLVSSGPYLFKFVS  123

Qy  182  GSHISRGFLLTY-----------ASS-------------DHPDLITC------  205
          ::                      :::              :|:
Db  124  DYETHGAGFSIRYEVFRKRGPECSRNFTSSSGMIKSPGFPEKYPNSLECTYIIFAPKMSEI  183

Qy  206  -LERASHYLKTEYSKFCPAGCRDVAGDIGSGNMVDGYRDT---SLLCKAAIHAGIIADEL  260
           :|      |::       | :|  : :|  :  :  :|::    :|  ::  |:|:|:
Db  184  ILEFESFELEPDSNTPGGARCRYDRLEI----NDGPVGHIGRYC-GGNFPGVRVSST  238

Qy  261  GQGISVLQRKGISRVEGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFEP----DGQIRA  316
          |  :|                   :|   :|::::|: |::|  ::|  :      ::| :
Db  239  GILSMVFYTDSAIAXKGFSANIVSVSQSSVSDEFQ-------CMEPLGMESGEIHSDQITV  291

Qy  317  SSSWQSVNESGDQVHWSPGQARLQDQGPSWASCDSSNNHKPRENLEIDLGERKKITGIRT  376
          |||  ||     :                | |||:|:||  :  ::  |:|:  ::|:|::
Db  292  SSQYSAI------------WSSERSRLNTPENGWTPGEDS-----VREHIQVDLGLLRFVSGIGT  339

Qy  377  TG--STQSNEFVYKSFVPWFMKNNNSKWKTYKGIVNNEEKVPQGNSNFRDPVQNNFIPPI  434
          ||    | |||: :|:            : | : : :||| |:|||| ||:|:|  | | :
Db  340  QQAISKETKKEYYLKTYRVDVSSNGEDWITLK--EGNKPVVFQGNSNPFDVVYRPFPKPV  397

Qy  435  VARVVRVRPQTWHGRIALRLVELIGCQIT  462
          : |:||:||::|  :::|||  :: ||:|:
Db  398  LTREVRIRPVSWENGVSLRFEVYGCKII  425
```

ID   NRP1_HUMAN   STANDARD;   PRT;   923 AA.
AC   O14786; O60461;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neuropilin-1 precursor (Vascular endothelial cell growth factor 165
DE   receptor).
GN   NRP1 OR NRP OR VEGF165R.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;

RN   [1]
RP   SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM).
RX   MEDLINE=97433084; PubMed=9287753;
RA   He Z., Tessier-Lavigne M.;
RT   "Neuropilin is a receptor for the axonal chemorepellent, semaphorin
RT   III.";
RL   Cell 90:739-751(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (MEMBRANE-BOUND ISOFORM), AND SEQUENCE OF 22-39.
RX   MEDLINE=98180099; PubMed=9529250;
RA   Soker S., Takashima S., Miao H.-Q., Neufeld G., Klagsbrun M.;
RT   "Neuropilin-1 is expressed by endothelial and tumor cells as an
RT   isoform-specific receptor for vascular endothelial growth factor.";
RL   Cell 92:735-745(1998).
RN   [3]
RP   SEQUENCE FROM N.A. (SOLUBLE/SNRP1 ISOFORM), AND SEQUENCE OF 22-31.
RC   TISSUE=Prostatic adenocarcinoma;
RX   MEDLINE=20183929; PubMed=10688800;
RA   Gagnon M.L., Bielenberg D.R., Gechtman Z., Miao H.-Q., Takashima S.,
RA   Soker S., Klagsbrun M.;
RT   "Identification of a natural soluble neuropilin-1 that binds vascular
RT   endothelial growth factor: In vivo expression and antitumor
RT   activity.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:2573-2578(2000).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=20303748; PubMed=10748121;
RA   Gluzman-Poltorak Z., Cohen T., Herzog Y., Neufeld G.;
RT   "Neuropilin-2 and neuropilin-1 are receptors for the 165-amino acid
RT   form of vascular endothelial growth factor (VEGF) and of placenta
RT   growth factor but only neuropilin-2 functions as a receptor for
RT   the 145-amino acid form of VEGF.";
RL   J. Biol. Chem. 275:18040-18045(2000).

CC   -!- FUNCTION: THE MEMBRANE-BOUND ISOFORM IS A RECEPTOR INVOLVED IN THE
CC   DEVELOPMENT OF THE CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE
CC   FORMATION OF CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS
CC   OUTSIDE THE NERVOUS SYSTEM. IT MEDIATES THE CHEMOREPULSANT
CC   ACTIVITY OF SEMAPHORINS. IT BINDS TO SEMAPHORIN 3A, THE PLGF-2
CC   ISOFORM OF PGF, THE VEGF-165 ISOFORM OF VEGF AND VEGF-B.
CC   COEXPRESSION WITH KDR RESULTS IN INCREASED VEGF-165 BINDING TO KDR
CC   AS WELL AS INCREASED CHEMOTAXIS. IT MAY REGULATE VEGF-INDUCED
CC   ANGIOGENESIS.
CC   -!- FUNCTION: THE SOLUBLE/SNRP1 ISOFORM BINDS VEGF-165 AND APPEARS
CC   TO INHIBIT ITS BINDING TO CELLS. IT CAN ALSO INDUCE APOPTOSIS BY
CC   SEQUESTERING VEGF-165. MAY BIND AS WELL VARIOUS MEMBERS OF THE
CC   SEMAPHORIN FAMILY. ITS EXPRESSION HAS AN AVERSE EFFECT ON BLOOD
CC   VESSEL NUMBER AND INTEGRITY.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. THE SOLUBLE/SNRP1
CC   ISOFORM IS SECRETED.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: MEMBRANE-BOUND (SHOWN HERE)
CC   AND SOLUBLE/SNRP1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: THE EXPRESSION OF MEMBRANE-BOUND AND
CC   SOLUBLE/SNRP1 ISOFORMS DOES NOT SEEM TO OVERLAP. THE MEMBRANE-
CC   BOUND ISOFORM IS EXPRESSED BY THE BLOOD VESSELS OF DIFFERENT
CC   TISSUES. IN THE DEVELOPING EMBRYO IT IS FOUND PREDOMINANTLY IN THE
CC   NERVOUS SYSTEM. IN ADULT TISSUES, IT IS HIGHLY EXPRESSED IN HEART
CC   AND PLACENTA; MODERATELY IN LUNG, LIVER, SKELETAL MUSCLE, KIDNEY
CC   AND PANCREAS; AND LOW IN ADULT BRAIN. THE SOLUBLE/SNRP1 ISOFORM IS
CC   FOUND IN LIVER HEPATOCYTES, KIDNEY DISTAL AND PROXIMAL TUBULES.

DR   EMBL; AF018956; AAC51759.1; -.
DR   EMBL; AF016050; AAC12921.1; -.
DR   EMBL; AF145712; AAF44344.1; -.
DR   MIM; 602069; -.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR001092; HLH_dim.
DR   InterPro; IPR000998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   Pfam; PF00629; MAM; 2.
DR   PRINTS; PR00020; MANDOMAIN.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS50060; MAM_2; 1.
KW   Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor;
KW   Alternative splicing.
FT   SIGNAL   1   21
FT   CHAIN   22   923   NEUROPILIN-1.
FT   DOMAIN   22   856   EXTRACELLULAR (POTENTIAL).

```
RESULT     6

ID    NRP1_RAT        STANDARD;      PRT;    922 AA.
AC    Q9QWJ9;
DT    16-OCT-2001 (Rel. 40, Created)
DT    16-OCT-2001 (Rel. 40, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    Neuropilin-1 precursor (Vascular endothelial cell growth factor 165
DE    receptor).
GN    NRP1.
OS    Rattus norvegicus (Rat).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX    NCBI_TaxID=10116;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=Sprague-Dawley;
RX    MEDLINE=97433085; PubMed=9287754;
RA    Kolodkin A.L., Levengood D.V., Rowe E.G., Tai Y.-T., Giger R.J.,
RA    Ginty D.D.;
RL    "Neuropilin is a semaphorin III receptor.";
RL    Cell 90:753-762(1997).
CC    -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC        CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC        CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
CC        SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
CC        BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165
CC        ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN
CC        INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
CC        IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
CC    -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC    -!- TISSUE SPECIFICITY: FOUND IN THE EMBRYONIC NERVOUS SYSTEM.
CC    -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC    -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC    -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC    -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; AF016296; AAC53337.1; -.
DR    InterPro; IPR000859; CUB.
DR    InterPro; IPR000421; FA58_C.
DR    InterPro; IPR001092; HLH_dim.
DR    InterPro; IPR000998; MAM.
DR    Pfam; PF00431; CUB; 2.
DR    Pfam; PF00754; F5_F8_type_C; 2.
DR    Pfam; PF00629; MAM; 1.
DR    PRINTS; PR00020; MAMDOMAIN.
DR    SMART; SM00042; CUB; 2.
DR    SMART; SM00231; FA58C; 2.
DR    SMART; SM00137; MAM; 1.
DR    PROSITE; PS01180; CUB; 2.
DR    PROSITE; PS01285; FA58C_1; 2.
DR    PROSITE; PS01286; FA58C_2; 2.
DR    PROSITE; PS00740; MAM_1; 1.
DR    PROSITE; PS50060; MAM_2; 1.
KW    Transmembrane; Glycoprotein; Neurone; Signal; Repeat; Receptor.
FT    SIGNAL         1     21
FT    CHAIN         22    922       NEUROPILIN-1.
FT    DOMAIN        22    855       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM     856    880       POTENTIAL.
FT    DOMAIN       881    922       CYTOPLASMIC (POTENTIAL).
FT    DOMAIN        27    141       CUB 1.
FT    DOMAIN       147    265       CUB 2.
FT    DOMAIN       275    424       F5/8 TYPE C 1.
FT    DOMAIN       431    583       F5/8 TYPE C 2.
FT    DOMAIN       645    811       MAM.
FT    DISULFID      27     54       PROBABLE.
FT    DISULFID      82    104       PROBABLE.
FT    DISULFID     147    173       PROBABLE.
FT    DOMAIN       645    811       MAM.
FT    DISULFID     206    228       BY SIMILARITY.
FT    DISULFID     275    424       BY SIMILARITY.
FT    DISULFID     431    583       BY SIMILARITY.
```

```
Query Match     124; Best Local Similarity   24.9%; Pred. No. 1.3e-17;
Matches  124; Conservative   76; Mismatches  176; Indels  122; Gaps  23;

Qy     64 RGL---LALLIAVSAPLRLQAEELGDGCHLVTYQDSGTMTSKNYPGTYPNHTVCEK 117
          |||   |:|:| :|:|                          |: : :||:|:|
Db      3 RGLPLCAVLALVLAPAGAFR---------NDKCGDPTIKIESPGYLTSPGYPHSYHPSKCEW 56

Qy    118 TIIVPKG-KRLILRLG-DLDIESQTCASDYL-LFTSSSDQ---YGPYCGSMTVPKELLLN 171
          |:||::| :|:||||  |:|  :||  : :|  |:  |||     | | | :||:||||
Db     57 LIQADPVPQRIMINFNFHPDLEDRDCKYDYVEVFDGENENGHRGKFCGKI-APPVVSS 115

Qy    172 TSEVTVRFESGSHISGRGFLLITYA------------SSDHPDLITC---205
          |  |:|| |: : |:: :|||   ||              |:|||:|:
Db    116 GPFLIKVSVDYETHGAGFSIRYEIFKRGPECSQNYTTPSGVIKSPGFPEKYPNSLECTY 175

Qy    206 ---------LERASHYLKIEYSKFCPAG--CRDVAGDISGNMVDGYRDTSLLCKAATH 252
                   |::|  | :|  |  ||      |:   |: :  |||  : : |:|  |
Db    176 IVFAPKMSEIILEFESFDLEPDSNP--PGGMFCRYDRLEI---VGPH 223

Qy    253 AG-IIADELGLGOI------------SVLQRKGISRYEGILANGVLSRDGSLSDKRFLF 297
          ||  :|| || : |            ||  :| :| ||:  |:|: : |:|   :|::
Db    224 IGRYCGOKTPGRIRSSSGLLSMVFYTLDSAIAKEGFSANYSVLQSSV-SEDFK 274

Qy    298 TSNGCSRSLSFEP----DCQIRASSMQSVNESDCQVHWSPGCQARLQDDCGPSWASGDSSN 353
          | :||  |:|::|    | |:  |: :| |  | ||: || ||: |::| | |:  |||:
Db    275 CMEALGMESGEIHSDSQITDASSOYSI-------NWSAERSRLNYPENGMTPGEDSY 322

Qy    354 NHKPREWLEIDLGEKKKITGLIRTTG---STQSNFNFEYVKSFVVMNFKNNSKWKIYKGIVNN 411
          |:|:: |  |||  ::|     |  |    :|::| :||  : : : | | | | : |:|
Db    323 ----RENIQVDLGLLRFVTAVGTQGAISKETRKKYYVKTKIDVSSNGEDWITIK--EGN 376

Qy    412 EERVFQGNSNFRDPVQNNFIPPIVARVRVRVPQTWHRQRIALKVELIGCOIT------OGNDS 467
          | |: |||:| ||||: | | |:| |::::| ::|| :|| :| :|  | |     |::|
Db    377 KPVLFQGNNTNPDVVVAVFPKRLITRVRVRIKPATWETGISMRFEVYGCKITDYPCSGMLG 436

Qy    468 LVHRKTSQSTSVSTKKED 485
          :|  :|   | |: |: |
Db    437 MVSGLLSDSQITSSNGQD 454
```

```
SEQUENCE   923 AA;  103120 MW;  ADEADCA4849ED557  CRC64;
```

```
RESULT 7

Qy  Match                                                                              922;

Qy  Best Local Similarity   10.8%;    Score 331;    DB 1;    Length 922;
    Matches 112;    Conservative 81;    Mismatches 166;    Indels 106;    Gaps 19;

SQ  SEQUENCE   922 AA;   103082 MW;   CC6F82AD09B0F2E CRC64;

Qy    64  RGI-------LALLLAVSAPLRQANERLADPGCGHLVTYQDSGTMTSKNFGCTFPNHTVCEK 117
Db     3  RGIPLLCATIALALALAGAFR------SDKCGGTFKIENPGYLTFSPGYPHSHPSEKCEW   56

Qy   118  TIIVPKG-KRILIRLG-DLDIESQCTCASDYLLFTSSSDQ---YGPYCCGSMTVFKELLIN 171
Db    57  LIQAEPPVQRIMIFNPHFDLEDRDCKYDIVEVIDGENEGGRLWGKPCCGKI-AFPSPVVSS 115

Qy   172  TSEVTVRFESGSHISGRGFLITYA----------------SSDHPDLITC-           205
Db   116  GPFLPIKPVSDYETHGAGSIRVEIFKRGPEDCSQNTAPTGVIKSPGFPEKYPNSLECTY   175

Qy   206  -------------LERASHYLKTEYSKFCPAG---CRDVAGDISGNMVDGYRDTSLLCKAAIH 252
Db   176  IFAPKMSEIILEFESFDLEQDSNP--PGGVFCYDRLEI--------WDGFPE------   219

Qy   253  AGTIADELGGOISVLQRKGISRVEGILA--------NGYLSRDGSLSDKRFLFTSNG-----CS 303
Db   220  --------VGPHIGRYCGQKTPGRTRSSSGLSMVFYTDSAIAKEGFSANYSVLGSSISEDFKCM 276

Qy   304  RSLSFEP---------DGQIRASSSKWQSVNEGSDQVHWSFCQQARLQDQGGPSWASGGSSNNHKPRR 359
Db   277  EALGHESGCEIHSDQTTASSQVGT----NWSVERSRLNYPENGWFPGEDSY---RE         324

Qy   360  WLEIDLGEKKIIGIRTTG--STQSNFNFYVKSFYWNNFKNNNSKWKTYKGIVNNEEKVFQ 417
Db   325  WIQVDLGLLRFVTAVGTQGAISKEFKKKYYVKTYRVDISSNGEDWITTLR--EGNKAITFQ 382

Qy   418  GNSNFRDPVQONFIPPIVARYVRAVVPQTWHQRIALKVELIGCCOTT         462
Db   383  GNTNPTLQVVFGVFKPLITRFVRIKPASWETGISMRFEVYGCKIT 427
```

```
FT   CARBOHYD   150    150   N-LINKED (GLCNAC...).   (POTENTIAL).
FT   CARBOHYD   261    261   N-LINKED (GLCNAC...).   (POTENTIAL).
FT   CARBOHYD   300    300   N-LINKED (GLCNAC...).   (POTENTIAL).
FT   CARBOHYD   522    522   N-LINKED (GLCNAC...).   (POTENTIAL).
FT   CARBOHYD   841    841   N-LINKED (GLCNAC...).   (POTENTIAL).
```

```
ID   FA5_BOVIN     STANDARD;     PRT;    2211 AA.
AC   Q28107; Q28108;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Coagulation factor V precursor (Activated protein C cofactor).
GN   F5.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92147638; PubMed=1737753;
RA   Guinto E.R., Esmon C.T., Mann K.G., Macgillivray R.T.;
RL   "The complete cDNA sequence of bovine coagulation factor V.";
RL   J. Biol. Chem. 267:2971-2978(1992).
CC   -!- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC       WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC   -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC       chain, noncovalently bound. The interaction between the two chains
CC       is calcium-dependent.
CC   -!- DOMAIN: Domain B contains 29.5 x 9 AA tandem repeats, and 2 X 14
CC       AA repeats.
CC   -!- PTM: Thrombin activates factor V proteolytically to the active
CC       cofactor factor Va (formation of a heavy chain and a light
CC       chain at the C-terminus).
CC   -!- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
CC       ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC       2 PLASTOCYANIN-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: STRONG; TO COAGULATION FACTOR VIII.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M81440; AAA30512.1; -.
DR   EMBL; M81441; AAA30513.1; -.
DR   HSSP; P00450; 1KCW.
DR   InterPro; IPR000117; Cu-oxidase.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW   Repeat.
FT   SIGNAL         1      28
FT   CHAIN         29    2211     POTENTIAL.
FT   CHAIN         29     741     COAGULATION FACTOR V.
FT   PEPTIDE      742    1564     HEAVY CHAIN (BY SIMILARITY).
FT   CHAIN                1453     ACTIVATION PEPTIDE (CONNECTING REGION)
FT   CHAIN       1565    2211                       (BY SIMILARITY).
FT   DOMAIN        30     196     LIGHT CHAIN (BY SIMILARITY).
FT   DOMAIN        30     327     F5/8 TYPE A 1.
FT   DOMAIN       197    1214     F5/8 TYPE A 1.
FT   REPEAT       206    1223                       2-3.
FT   REPEAT       215    1232                       2-4.
FT   REPEAT       224    1241                       2-5.
FT   REPEAT       233    1250                       2-6.
FT   DOMAIN       203    1259     PLASTOCYANIN-LIKE 1.
FT   DOMAIN       348    1268     PLASTOCYANIN-LIKE 2.
FT   REPEAT       348     686                       2-7.
FT   DOMAIN       525     686     F5/8 TYPE A 2.
FT   DOMAIN       535    1277     PLASTOCYANIN-LIKE 3.
FT   DOMAIN       696    1286     PLASTOCYANIN-LIKE 4.
FT   SIMILAR      899    1295     B.
FT   DOMAIN      1124    1304     30 X 9 AA REPEATS IN HUMAN FA5.
FT   REPEAT      1124    1151                       1-1.
FT   REPEAT      1137    1313                       1-2.
FT   REPEAT      1138    1322     2 X 14 AA TANDEM REPEATS.
FT   DOMAIN      1188    1453     ACTIVATION PEPTIDE (CONNECTING REGION)
FT                             30 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT                             [AS]-L-S-P-D-[LP]-[GS]-Q-[TE].
FT   REPEAT      1188    1196                       2-1.
FT   REPEAT      1197    1205                       2-2.
FT   REPEAT      1206    1214                       2-3.
FT   REPEAT      1215    1223                       2-4.
FT   REPEAT      1224    1232                       2-5.
FT   REPEAT      1233    1241                       2-6.
FT   REPEAT      1242    1250                       2-7.
FT   REPEAT      1251    1259                       2-8.
FT   REPEAT      1260    1268                       2-9.
FT   REPEAT      1269    1277                       2-10.
FT   REPEAT      1278    1286                       2-11.
FT   REPEAT      1287    1295                       2-12.
FT   REPEAT      1296    1304                       2-13.
FT   REPEAT      1305    1313                       2-14.
FT   REPEAT      1314    1322                       2-15.
FT   REPEAT      1323    1331                       2-16.
FT   REPEAT      1332    1340                       2-17.
FT   REPEAT      1341    1349                       2-18.
FT   REPEAT      1350    1358                       2-19.
FT   REPEAT      1359    1367                       2-20.
FT   REPEAT      1368    1376                       2-21.
FT   REPEAT      1377    1385                       2-22.
FT   REPEAT      1386    1394                       2-23.
```

```
Qy    371  ITGIRTGTSTQOSNFNFPYVKSEFVNNFKNNNSKWRTYKGITVNNEEKVFQGNSNFRDPVQGNNF  430
           |||| :|| | |||||  |:|   | ||  ||:|||||| |:|| |::|:|| |  ||:|
Db   2120  ITAIVTGCKSLSSEMVYVKSYTIHYSDGGTDWKPYREKSSMVDKIFEGNNNVRGHVKNFF   2179

Qy    431  IPPIVANYVRVPQTMHQRIALKVELGIGCCI   461
           || :| |||| | ||::| |||:|| :||||:
Db   2180  NPPIISRFIIIPFKTWNQSIALRELFQGCDM   2210
```

RESULT 8

```
NRP1_XENLA     STANDARD;      PRT;    928 AA.
AC   P28824;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neuropilin-1 precursor (A5 protein) (A5 antigen).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=91137458; PubMed=1908252;
RA   Takagi S., Hirata T., Agata K., Mochii M., Eguchi G., Fujisawa H.;
RT   "The A5 antigen, a candidate for the neuronal recognition molecule,
RT   has homologies to complement components and coagulation factors.";
RL   Neuron 7:295-307(1991).
CC   -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC       CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC       CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
CC       SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS (BY
CC       SIMILARITY). PRESUMED TO BE INVOLVED IN THE NEURONAL RECOGNITION
CC       BETWEEN THE OPTIC NERVE FIBERS AND THE VISUAL CENTERS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: RETINAL GANGLION CELLS AND VISUAL CENTER
CC       NEURONS.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
```

```
DR   EMBL; D10467; BAA01260.1; -.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; F58_C.
DR   InterPro; IPR000998; MAM.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
DR   PROSITE; PS50060; MAM_2; 1.
KW   Antigen; Glycoprotein; Repeat; Receptor;
KW   Transmembrane; Signal; Neurone.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    928       NEUROPILIN-1.
```

```
Query Match             10.8%;  Score 331;  DB 1;  Length 2211;
Best Local Similarity   32.5%;  Pred. No. 9.2e-17;
Matches   88;  Conservative  48;  Mismatches  89;  Indels  46;  Gaps  10;
```

```
Qy    209  ASHYLKTEY-SKFCPA------GCRDVAGDISGNNV--DGYRDTSLLCKAAIHAGIADE    259
           : :|| : :  : || |        || ||  |||:    || | |  |  | || |||:
Db   1968  AKHYLKPYTTEPCVAYSLDRKNWRIFKGNSTRNYWYFGGNSDASIVIKENGIQPPVVA--  2025

Qy    260  LGGQISVLQRKGISRYEGILANGVLSRDGSLSDKRFLTSNGCSRSLSFE----PDGQIR    315
           |      |: | |       ||  |:  |   |  | ||    :|   |      |  |:
Db   2026  ----RYTRISPTGSYNKPALRLELQGC-EVNGCSTPLGMESGKIENKOTT        2070

Qy    316  ASS----SWQSVNESGDQVHWSPGQARLQDQG---PSMASGDSSNNHKPREWLEIDLGEKKN  370
           |||      | || |         | ||| |     |  :|| : |    |   |:||| : :
Db   2071  ASSFKKSWWG------------NYKEPPLARLNAQGRVNAMQAKANNNN----OWLQIDLLKIRK  2119
```

```
FT   REPEAT      1395   1403       2-24.
FT   REPEAT      1404   1412       2-25.
FT   REPEAT      1413   1421       2-26.
FT   REPEAT      1422   1430       2-27.
FT   REPEAT      1431   1439       2-28.
FT   REPEAT      1440   1444       2-29. (PARTIAL).
FT   REPEAT      1445   1453       2-30.
FT   DISULFID    1569   1890
FT   DISULFID    1569   1890
FT   DOMAIN      1738   1890       PLASTOCYANIN-LIKE 5.
FT   DOMAIN      1748   1890       PLASTOCYANIN-LIKE 6.
FT   DOMAIN      1894   2048       F5/8 TYPE C 1.
FT   DOMAIN      2053   2208       F5/8 TYPE C 2.
FT   SITE         741    742       PLASTOCYANIN-LIKE A 3.
FT   SITE        1034   1035       CLEAVAGE (BY THROMBIN).
FT   SITE        1564   1565       CLEAVAGE (BY THROMBIN).
FT   DISULFID     167    193       CLEAVAGE (BY SIMILARITY).
FT   DISULFID     499    525       PROBABLE.
FT   DISULFID    1712   1738       PROBABLE.
FT   DISULFID    1894   2048       PROBABLE.
FT   DISULFID    2053   2208       PROBABLE.
FT   MOD_RES      697    697       SULFATION (POTENTIAL).
FT   MOD_RES      701    701       SULFATION (POTENTIAL).
FT   MOD_RES      730    730       SULFATION (POTENTIAL).
FT   MOD_RES     1513   1513       SULFATION (POTENTIAL).
FT   MOD_RES     1529   1529       SULFATION (POTENTIAL).
FT   MOD_RES     1537   1537       SULFATION (POTENTIAL).
FT   MOD_RES     1541   1541       SULFATION (POTENTIAL).
FT   CARBOHYD     225    225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     239    239       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     297    297       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     382    382       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     460    460       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     553    553       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     587    587       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     745    745       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     756    756       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     774    774       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     780    780       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     902    902       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     952    952       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     964    964       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1044   1044       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1053   1053       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1062   1062       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1071   1071       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1078   1078       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1094   1094       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1451   1451       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1490   1490       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1550   1550       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1690   1690       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1839   1839       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    1997   1997       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    2196   2196       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT      587    592       NFTLFA -> T (IN VARIANT 2).
SQ   SEQUENCE   2211 AA;  248981 MW;  CBBF9287386670C45 CRC64;
```

Left column

```
FT   DOMAIN       22    860   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    861    883   POTENTIAL.
FT   DOMAIN      884    928   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       27    141   CUB 2.
FT   DOMAIN      147    265   CUB 2.
FT   DOMAIN      275    424   F5/8 TYPE C 1.
FT   DOMAIN      431    584   F5/8 TYPE C 2.
FT   DOMAIN      646    812   MAM.
FT   DISULFID     27     54   PROBABLE.
FT   DISULFID     82    104   PROBABLE.
FT   DISULFID    147    173   PROBABLE.
FT   DISULFID    206    228   PROBABLE.
FT   DISULFID    275    424   BY SIMILARITY.
FT   DISULFID    431    584   BY SIMILARITY.
FT   CARBOHYD    150    150   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    261    261   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    300    300   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    523    523   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    844    844   N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE    928 AA;  103416 MW;  AF68323B0A4C789D CRC64;
```

Query Match
Best Local Similarity: 10.8%;   Score 330;   DB 1;   Length 928;
Matches 112;   Conservative 74;   Mismatches 171;   Indels 106;   Gaps 20;
Similarity 24.2%;   Pred. No. 3.2e-17;

```
Qy   66  LLALLLAVSAPLRLQAEELGDGCGHLVTYQDSGTMTSKNYPGTYPNTTVCEKTITVPKG-  124
Db    5  LLSCCCWLLGSLRSSWASRNDKCGDITIKITSPSYLTSAGYPHSYTPPSQRCEWLLQAPEHY  64

Qy  125  KRLILRLG-DLDTESQTCASDYLLF----TSSSDQYGPYCGSMTVPKELILNTSEVTVRF  179
Db   65  QRIMINFNHFDLEDRECKYDYVEYIDGDNANGQLLGKYCGKI-APSPLVSYGPSIFIFRF  123

Qy  180  ESGSHISGRGFLTY------ASSD------HPDLIYC------  205
Db  124  VSDYETFGAGFSIRYEVFKTGPECSRNFTSSNGVIKSPKYPEKYPNALECTYIIFARKMQ  183

Qy  206  ---LERASHYLRKTEYSKPCPAG--CR-DVAGDISGNMVDGYRDFSLLCKAAIHAGIIADE  259
Db  184  BIVLEFESPELEADSN--APGGQTCRYDWLG-----IWDGFPG-----VGPH  223

Qy  260  IGRYCGQNTPGRVRSFTGILSMIFHTDSAIAKEGFFANFSVVQSNTDEDF-----QCKEA  278
Db  224  NGVLSRDG-----SLSDKRFLFTSNGCSRS  305

Qy  306  LSFEPD-----GQIRASSWQSVNESGDQVHWSGQARLQDQGPSWASGDSSNNHKPREWL  361
Db  279  LGMESGEIHFDQISVSSQY----SMWNSAERSRLNVYENGWTPGEDT---VKEWI  326

Qy  362  EIDLGEKKKITGIIRFTG--STQSNFPYVVKSFVMNFKNNNSKWKTYKGIVNNEEKVPQGN  419
Db  327  QVDLENLRFVSGIGTQGA-ISKETKKKYPVKSYKVDISSNGEDWITILKD--GNKHLVFTGN  384

Qy  420  SNFRDPVQNNFIPPIVARYVRVFPQTHQRIALKVELLCGCIT  462
Db  385  TDATDVVYRPFSKPVITRFVRLRPVTWENGISLRFELYGCKIT  427
```

Right column

RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Embryonic brain;
RX   MEDLINE=96353149; PubMed=8748368;
RA   Kawakami A., Kitsukawa T., Takagi S., Fujisawa H.;
RT   "Developmentally regulated expression of a cell surface protein,
RT   neuropilin, in the mouse nervous system.";
RL   J. Neurobiol. 29:1-17(1996).

```
CC   -!- FUNCTION: RECEPTOR INVOLVED IN THE DEVELOPMENT OF THE
CC       CARDIOVASCULAR SYSTEM, IN ANGIOGENESIS, IN THE FORMATION OF
CC       CERTAIN NEURONAL CIRCUITS AND IN ORGANOGENESIS OUTSIDE THE NERVOUS
CC       SYSTEM. IT MEDIATES THE CHEMOREPULSANT ACTIVITY OF SEMAPHORINS. IT
CC       BINDS TO SEMAPHORIN 3A, THE PLGF-2 ISOFORM OF PGF, THE VEGF-165
CC       ISOFORM OF VEGF AND VEGF-B. COEXPRESSION WITH KDR RESULTS IN
CC       INCREASED VEGF-165 BINDING TO KDR AS WELL AS INCREASED CHEMOTAXIS.
CC       IT MAY REGULATE VEGF-INDUCED ANGIOGENESIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: NERVOUS SYSTEM.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
```

```
DR   EMBL; D50086; BAA08789.1; -.
DR   MGD; MGI:106206; Nrp.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR008050; NRP.
DR   InterPro; IPR000421; F58.
DR   InterPro; IPR000998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00754; F5_8_type_C; 2.
DR   Pfam; PF00629; MAM; 1.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS50060; MAM_2; 1.
KW   Transmembrane; Glycoprotein; Receptor.
FT   SIGNAL        1     21
FT   CHAIN        22    923   NEUROPILIN-1.
FT   DOMAIN       22    856   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    857    879   POTENTIAL.
FT   DOMAIN      880    923   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       27    141   CUB 1.
FT   DOMAIN      147    265   CUB 2.
FT   DOMAIN      275    424   F5/8 TYPE C 1.
FT   DOMAIN      431    583   F5/8 TYPE C 2.
FT   DOMAIN      645    811   MAM.
FT   DISULFID     27     54   PROBABLE.
FT   DISULFID     82    104   PROBABLE.
FT   DISULFID    147    173   PROBABLE.
FT   DISULFID    206    228   PROBABLE.
FT   DISULFID    275    424   BY SIMILARITY.
FT   DISULFID    431    583   BY SIMILARITY.
FT   CARBOHYD    150    150   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    261    261   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    300    300   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    522    522   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    842    842   N-LINKED (GLCNAC...)  (POTENTIAL).
SQ   SEQUENCE    923 AA;  103020 MW;  0644B8A1707965808 CRC64;
```

Query Match
Best Local Similarity: 10.8%;   Score 329;   DB 1;   Length 923;
Similarity 24.0%;   Pred. No. 3.7e-17;

RESULT 9
NRP1_MOUSE

```
ID   NRP1_MOUSE     STANDARD;      PRT;   923 AA.
AC   P97333;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neuropilin-1 precursor (A5 protein).
GN   NRP1 OR NRP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
```

Matches 113; Conservative 79; Mismatches 161; Indels 118; Gaps 21;

```
Qy     64  RGL-------LALILAVSAPLRLQAEELGDGCGHLVTVQDSGTMSKNYPGTYPNHTVCEK 117
               |||       :||:|  :|:   |    |:| :|::| | | ::|   | | |  |:|
Db      3  RGLPLLCATLAIALALALAGAFR-----SDKCGGGTFIKTENPGVITSPGYPHSYHPSEKCEW  56

Qy    118  TITVPKG-KRLILRLG-DLDIESOTCASDYLLFSSSDQ-----YGPYCGSMTVPKEILN 171
               ||||| |:|||  |  |:|| | || | | | :|||         | || | :|:|  |
Db     57  LIQAPEPYQRIIINFNFHFDLEDRDCKYDYVEVIDGENESGRLWGKFCGKI-APSPVSS 115

Qy    172  TSEVTVRFBSGSHISGKGFHLTYA----------------------SSDHPDLITC--- 205
               |:|   : || |   :|   | |                        || || |
Db    116  GPFLFIKFVSDYETHGAGFSIRYEIFKRGPECSQNYTAPTGVIKSPGFPEKYPNCIECTY 175

Qy    206  -----LERASHYLKTEYSKFCCPAG---CRDVAGDISGNNVDGYRDTSLLCKAAIH 252
                       |:||  :|   |       |: || |  | :| | |:| || | |
Db    176  IIFAPKMSEIILEFESFDLEQDSNP--PGGMFCRVDRLEI----WDGFPE----VGPH 223

Qy    253  AG-IIADELGGOI------------SVLQRKGISRYEGILANGVLSRPGSLSDKRFLF 297
               || :|:| ||                ||  |:| ||:| | |  | |: |  |  |
Db    224  IGRYCGQKTPGRIRSSGVLSMVFYTDSAIAKEGFSANYSLVLQSSI-SEDFK------ 274

Qy    298  TSNGCSRSLSFEP----DGQIRASSSWQSVNESGDQVHWSPGGQARLQDOGPSNASGDSSN 353
               |:| | ||  :|     ||  || |  |  || | ||  |:    | |  ::| |||  |
Db    275  -----CMEALGMEGGEIHSDQITASSQYGT------NWSVERSRLNTPENGWTPGEDSY 322

Qy    354  NHKPREWLEIDLGEKKIIGIRTLG---STQSNFNFYVKSFVVMKFKNNNSKMKTYKGIVNN 411
               |:| | |||  :||  |||  |       |  ||| :| :| |  |:|  | :| :| |
Db    323  ----KEWIQVDLGLLREYTVAGYTQGAISKEPKKKYYVKTYRDISSNGEDWISLK--EGN 376

Qy    412  EEKVFQGNSNFRDPVQNRFIPIVARYVRNVPQTWHQRIALKVELICQQIT 462
               :|| ||| ||     |  |   ||||||    |:|  | :|  :|| :|
Db    377  KAIIFQGNTNPTDVVLGVFSKPLITRFVRIKPVSWETGISMRPEVYGCKIT 427
```

RESULT 10

```
FA5_HUMAN
ID    FA5_HUMAN            STANDARD;      PRT;    2224 AA.
AC    P12259; Q14285;
DT    01-OCT-1989 (Rel. 12, Created)
DT    01-FEB-1994 (Rel. 28, Last sequence update)
DT    16-OCT-2001 (Rel. 40, Last annotation update)
DE    Coagulation factor V precursor (Activated protein C cofactor).
GN    F5.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=92332668; PubMed=1567832;
RA    Cripe L.D., Moore D.D., Kane W.H.;
RT    "Structure of the gene for human coagulation factor V.";
RL    Biochemistry 31:3777-3785(1992).
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=87260886; PubMed=3110773;
RA    Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
RA    Hewick R.M., Kaufman R.J., Mann K.G.;
RT    "Complete cDNA and derived amino acid sequence of human factor V.";
RL    Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
RN    [3]
RP    SEQUENCE OF 1-1600 FROM N.A.
RX    MEDLINE=88107560; PubMed=2827731;
RA    Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
RT    "Cloning of cDNAs coding for the heavy chain region and connecting
RT    region of human factor V, a blood coagulation factor with four types
RT    of internal repeats.";
RL    Biochemistry 26:6508-6514(1987).
RN    [4]
RP    SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
RX    MEDLINE=86317665; PubMed=3092220;
RA    Kane W.H., Davie E.W.;
```

```
RT    "Cloning of a cDNA coding for human factor V, a blood coagulation
RT    factor homologous to factor VIII and ceruloplasmin.";
RL    Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
RN    [5]
RP    PARTIAL SEQUENCE FROM N.A.
RC    TISSUE=Fibroblast;
RX    MEDLINE=93205619; PubMed=8454869;
RA    Shen N.L., Fan S.-T., Pyati J., Graff R., Lapolla R.J.,
RA    Edgington T.S.;
RT    "The serine protease cofactor factor V is synthesized by
RT    lymphocytes.";
RL    J. Immunol. 150:2992-3001(1993).
RN    [6]
RP    SULFATION.
RX    MEDLINE=94264012; PubMed=8206629;
RA    Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
RA    Kaufman R.J.;
RT    "Posttranslational sulfation of factor V is required for efficient
RT    thrombin cleavage and activation and for full procoagulant activity.";
RL    Biochemistry 33:6952-6959(1994).
RN    [7]
RP    SULFATION.
RX    MEDLINE=90366699; PubMed=2168225;
RA    Hortin G.L.;
RT    "Sulfation of tyrosine residues in coagulation factor V.";
RL    Blood 76:946-952(1990).
RN    [8]
RP    X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
RX    MEDLINE=94217810; PubMed=8164741;
RA    Bertina R.M., Koeleman B.P.C., Koster T., Rosendaal F.R.,
RA    Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
RT    "Mutation in blood coagulation factor V associated with resistance to
RT    activated protein C.";
RL    Nature 369:64-67(1994).
RN    [9]
RP    VARIANT APCR GLN-534.
RX    MEDLINE=94217810; PubMed=8164741;
RA    Macedo-Ribeiro S., Bode W., Huber R., Quinn-Allen M.A., Kim S.W.,
RA    Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
RA    Fuentes-Prior P.;
RT    "Crystal structures of the membrane-binding C2 domain of human
RT    coagulation factor V.";
RL    Nature 402:434-439(1999).
CC   -!- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC        WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC   -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC        chain, noncovalently bound. The interaction between the two chains
CC        is calcium-dependent.
CC   -!- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
CC        REPEATS.
CC   -!- PTM: Thrombin activates factor V proteolytically to the active
CC        cofactor, factor Va (formation of a heavy chain at the N-
CC        terminus and a light chain at the C-terminus).
CC   -!- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
CC        ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY.
CC   -!- DISEASE: OWREN PARAHEMOPHILIA, AN HEMORRHAGIC DIATHESIS, IS DUE
CC        TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
CC        FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR); THE APCR
CC        MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
CC        A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
CC        IMPLANTATION.
CC   -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC        2 PLASTOCYANIN-LIKE REPEATS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
```

```
CC   EMBL; L32779; AAB59401.1; -.
DR   EMBL; L32755; AAB59401.1; -.
DR   EMBL; L32756; AAB59401.1; JOINED.
DR   EMBL; L32757; AAB59401.1; JOINED.
DR   EMBL; L32758; AAB59401.1; JOINED.
DR   EMBL; L32759; AAB59401.1; JOINED.
DR   EMBL; L32760; AAB59401.1; JOINED.
DR   EMBL; L32761; AAB59401.1; JOINED.
DR   EMBL; L32762; AAB59401.1; JOINED.
DR   EMBL; L32763; AAB59401.1; JOINED.
DR   EMBL; L32764; AAB59401.1; JOINED.
DR   EMBL; L32765; AAB59401.1; JOINED.
DR   EMBL; L32766; AAB59401.1; JOINED.
DR   EMBL; L32767; AAB59401.1; JOINED.
DR   EMBL; L32768; AAB59401.1; JOINED.
DR   EMBL; L32769; AAB59401.1; JOINED.
DR   EMBL; M14335; AAB59401.1; -.
DR   EMBL; M16967; AAA52424.1; -.
DR   EMBL; L32770; AAB59532.1; -.
DR   EMBL; L32771; AAB59532.1; JOINED.
DR   EMBL; L32772; AAB59532.1; JOINED.
DR   EMBL; L32773; AAB59532.1; JOINED.
DR   EMBL; L32774; AAB59532.1; JOINED.
DR   EMBL; L32765; AAB59532.1; JOINED.
DR   EMBL; L32767; AAB59532.1; JOINED.
DR   EMBL; L32766; AAB59532.1; JOINED.
DR   EMBL; L32777; AAB59532.1; JOINED.
DR   EMBL; L32778; AAB59532.1; JOINED.
DR   EMBL; L32769; AAB59532.1; JOINED.
DR   PIR; A25897; A25897.
DR   PIR; A28028; A28028.
DR   PDB; 1CZS; 26-NOV-99.
DR   PDB; 1CZT; 26-NOV-99.
DR   PDB; 1CZV; 26-NOV-99.
DR   MIM; 134400; -.
DR   MIM; 188055; -.
DR   MIM; 227400; -.
DR   MIM; 227310; -.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Blood coagulation; Glycoprotein; Calcium; Signal; Zymogen;
KW   Repeat; Polymorphism; Disease mutation; Thrombophilia; 3d-structure.
FT   SIGNAL        1     28
FT   CHAIN        29   2224       COAGULATION FACTOR V.
FT   CHAIN        29    737       HEAVY CHAIN.
FT   PEPTIDE     738   1573       ACTIVATION PEPTIDE (CONNECTING REGION).
FT   CHAIN      1574   2224       LIGHT CHAIN.
FT   DOMAIN       30    329       F5/8 TYPE A 1.
FT   DOMAIN       30    193       PLASTOCYANIN-LIKE 1.
FT   DOMAIN      203    329       PLASTOCYANIN-LIKE 2.
FT   DOMAIN      348    684       F5/8 TYPE A 2.
FT   DOMAIN      348    526       PLASTOCYANIN-LIKE 3.
FT   DOMAIN      536    684       PLASTOCYANIN-LIKE 4.
FT   DOMAIN      692   1573       B.
FT   REPEAT      895    928       1-1.
FT   REPEAT      895    911       1-1.
FT   REPEAT      912    928       1-2.
FT   SIMILAR    1135   1148       2 X 17 AA TANDEM REPEATS.
FT   DOMAIN     1185   1501
FT   SIMILAR    1135   1501       35 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT                                [TNP]-L-S-P-D-L-S-Q-T.
FT   REPEAT     1185   1193       2-1.
FT   REPEAT     1194   1202       2-2.
FT   REPEAT     1203   1211       2-2.
FT   REPEAT     1212   1220       2-4.
FT   REPEAT     1221   1229       2-5.
FT   REPEAT     1230   1238       2-6.
FT   REPEAT     1239   1247       2-7.
FT   REPEAT     1248   1256       2-8.
FT   REPEAT     1257   1265       2-9.
FT   REPEAT     1266   1274       2-10.
FT   REPEAT     1275   1283       2-11.
FT   REPEAT     1284   1292       2-12.
FT   REPEAT     1293   1301       2-13.
FT   REPEAT     1302   1310       2-14.
FT   REPEAT     1311   1319       2-15.
FT   REPEAT     1320   1328       2-16.
FT   REPEAT     1329   1337       2-17.
FT   REPEAT     1338   1346       2-18.
FT   REPEAT     1347   1355       2-19.
FT   REPEAT     1356   1364       2-20.
FT   REPEAT     1365   1373       2-21.
FT   REPEAT     1374   1382       2-22.
FT   REPEAT     1383   1391       2-23.
FT   REPEAT     1392   1400       2-24.
FT   REPEAT     1401   1409       2-25.
FT   REPEAT     1410   1418       2-26.
```

```
RESULT 11

Query Match             10.4%;
Best Local Similarity   42.0%;
Matches  71;  Conservative  30;  Mismatches  52;  Indels  16;  Gaps   5;
Score 317;  DB 1;  Length 2224;  Pred. No. 1.1e-15;

Qy   300 NGCSRSLSFE----PDGQIRASS---SWQSVNESGDQVHWSPGQARLQDQGPSWASGDSS 352
         |||  |  |        |||  |    ||  |  |  | |  |    |   | |  |
Db  2064 NGCSTPLGMENGKIENKQITASSFKSWW----GD--YWEPFRARILNAQGRVNAWQAKA 2116

Qy   353 NNHKPREWLEIDLGEKKKITIGIRTTGSTNSNFPYVKSFVMNFKNNNSKWKTYKGIVNNE 412
         |||| |  |||||| | |  ||  |  |  || |  |    |   :     :|||| |
Db  2117 NNNK--QWLEIDLLKIKKITPALITQGCKSLSSEMVKSYTIHYSEQGVEWKPYVRLKSSMV 2174

Qy   413 EKVPGQNSNFRDPVQNNFIPPIVARYVRVYPQTHYRQRIALKVELLGCDI 461
         :|  |  |  | |  ||| |||| :  |: |  :| | |: |:||:|||
Db  2175 DKIFEGNTNTKGHVKNFPNPPIISRFIRVIPKTWNQSITLRLELFGCDI 2223
```

```
ID   FA5_PIG      STANDARD;      PRT;  2258 AA.
AC   Q9GLP1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Coagulation factor V precursor (Activated protein C
DE   cofactor).
GN   F5.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING OF F5/8 TYPE A AND C
RP   DOMAINS.
RC   TISSUE=Liver;
RX   MEDLINE=21121490; PubMed=11229814;
RA   Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,
RA   Kim H.K.W.;
RT   "Porcine factor V: cDNA cloning, gene mapping, three-dimensional
RT   protein modeling of membrane binding sites and comparative anatomy of
RT   domains.";
RL   Cell. Mol. Life Sci. 58:148-159(2001).
CC   -!- FUNCTION: Coagulation factor V is a cofactor that participates
CC       with factor Xa to activate prothrombin to thrombin.
CC   -!- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC       chain, noncovalently bound. The interaction between the two chains
CC       is calcium-dependent.
CC   -!- DOMAIN: Domain B contains 41 X 9 AA tandem repeats. Domains C1
CC       and C2 may be involved in membrane binding.
CC   -!- PTM: Thrombin activates factor V proteolytically to the active
CC       cofactor factor Va (formation of a heavy chain at the N-
CC       terminus and a light chain at the C-terminus).
CC   -!- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC       2 PLASTOCYANIN-LIKE REPEATS.
```

```
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -!- SIMILARITY: STRONG TO COAGULATION FACTOR VIII.
CC
```

```
DR   EMBL; AF191308; AAG28381.1; -.
DR   InterPro; IPR001117; Cu-oxidase.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00394; Cu-oxidase; 3.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00079; MULTICOPPER_OXIDASE1; 1.
KW   Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW   Repeat.
FT   SIGNAL        1     22
FT   CHAIN        23   2258       COAGULATION FACTOR V.
FT   CHAIN        23   2258       HEAVY CHAIN (BY SIMILARITY).
FT   PEPTIDE     738   1611       ACTIVATION PEPTIDE (CONNECTING REGION)
FT                                (BY SIMILARITY).
FT   CHAIN      1612   2258       LIGHT CHAIN (BY SIMILARITY).
FT   DOMAIN       30    329       F5/8 TYPE A 1.
FT   DOMAIN       30    193       PLASTOCYANIN-LIKE 1.
FT   DOMAIN      203    329       PLASTOCYANIN-LIKE 2.
FT   DOMAIN      348    683       F5/8 TYPE A 2.
FT   DOMAIN      348    525       PLASTOCYANIN-LIKE 3.
FT   DOMAIN      525    683       PLASTOCYANIN-LIKE 4.
FT   DOMAIN      691   1611       B.
FT   DOMAIN     1168   1539       41 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT                                T-L-S-P-D-L-[GS]-[HQ]-T.
FT   REPEAT     1168   1176       1.
FT   REPEAT     1177   1185       2.
FT   REPEAT     1186   1194       3.
FT   REPEAT     1195   1203       4.
FT   REPEAT     1204   1212       5.
FT   REPEAT     1213   1221       6.
FT   REPEAT     1222   1230       7.
FT   REPEAT     1231   1239       8.
FT   REPEAT     1240   1248       9.
FT   REPEAT     1249   1257       10.
FT   REPEAT     1258   1266       11.
FT   REPEAT     1267   1275       12.
FT   REPEAT     1276   1284       13.
FT   REPEAT     1285   1293       14.
FT   REPEAT     1294   1302       15.
FT   REPEAT     1303   1311       16.
FT   REPEAT     1312   1320       17.
FT   REPEAT     1321   1329       18.
FT   REPEAT     1330   1338       19.
FT   REPEAT     1339   1347       20.
FT   REPEAT     1348   1356       21.
FT   REPEAT     1357   1365       22.
FT   REPEAT     1366   1374       23.
FT   REPEAT     1375   1383       24.
FT   REPEAT     1384   1392       25.
FT   REPEAT     1393   1401       26.
FT   REPEAT     1402   1410       27.
FT   REPEAT     1411   1419       28.
FT   REPEAT     1420   1428       29.
FT   REPEAT     1429   1437       30.
FT   REPEAT     1438   1446       31.
FT   REPEAT     1447   1455       32.
FT   REPEAT     1456   1464       33.
FT   REPEAT     1465   1473       34.
FT   REPEAT     1474   1482       35.
FT   REPEAT     1483   1491       36.
FT   REPEAT     1492   1500       37.
FT   REPEAT     1501   1509       38.
FT   REPEAT     1510   1518       39.
FT   REPEAT     1519   1527       40.
FT   REPEAT     1531   1539       41.
FT   DOMAIN     1616   1941       F5/8 TYPE A 3.
FT   DOMAIN     1785   1941       PLASTOCYANIN-LIKE 5.
FT   DOMAIN     1795   1941       PLASTOCYANIN-LIKE 6.
FT   DOMAIN     1942   2095       F5/8 TYPE C 1.
FT   DOMAIN     2100   2255       F5/8 TYPE C 2.
FT   SITE        737    738       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   SITE       1029   1030       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   SITE       1611   1612       CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT   DISULFID    167    193       BY SIMILARITY.
FT   DISULFID    499    525       PROBABLE.
FT   DISULFID    725    745       PROBABLE.
FT   DISULFID   1759   1785       PROBABLE.
FT   DISULFID   1941   2095       BY SIMILARITY.
FT   DISULFID   2100   2255       BY SIMILARITY.
FT   MOD_RES     692    692       SULFATION (BY SIMILARITY).
FT   MOD_RES     696    696       SULFATION (BY SIMILARITY).
FT   MOD_RES     724    724       SULFATION (BY SIMILARITY).
FT   MOD_RES     726    726       SULFATION (BY SIMILARITY).
FT   MOD_RES     745    745       SULFATION (BY SIMILARITY).
FT   MOD_RES    1560   1560       SULFATION (POTENTIAL).
FT   MOD_RES    1576   1576       SULFATION (POTENTIAL).
FT   MOD_RES    1581   1581       SULFATION (POTENTIAL).
FT   MOD_RES    1584   1584       SULFATION (POTENTIAL).
FT   MOD_RES    1588   1588       SULFATION (POTENTIAL).
FT   MOD_RES    1631   1631       SULFATION (POTENTIAL).
FT   CARBOHYD    153    153       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    225    225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    297    297       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    382    382       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    460    460       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    467    467       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    553    553       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    741    741       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    752    752       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    760    760       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    776    776       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    782    782       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    899    899       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    960    960       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1048   1048       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1057   1057       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1066   1066       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1073   1073       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1089   1089       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1174   1174       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1480   1480       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1537   1537       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1597   1597       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1737   1737       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1886   1886       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   2044   2044       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   2243   2243       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   2258 AA;  256078 MW;  9159B9E0076A2ACC  CRC64;
```

```
Query Match      39.3%;  Score 314;  DB 1;  Length 2258;
Best Local Similarity   67;  Pred. No. 1.9e-15;
Matches    67;  Conservative   35;  Mismatches   49;  Indels   20;  Gaps   6;

Qy      300 NCCSRSLSFB------PDCGQIRASS--SWQSVNESGDQVHWSPGQARLQDQG--PSWASGD 350
                  ::  :  :       ::  :  :    : ||  :  ::  :     :  | |  :
Db     2098 NGCSTPLGMESGNIRNEDITASSFFKSWW-----GD--YWPFFRARLRNAOGRVNANQARA 2150

Qy      351 SSNHHKFPREMLEIDLGEKKKTGIRTTCSTQSNFNFYVXSEVNNFKNNNSKWKTYKGIVN 410
               : :|     ::  :||  : |||   :  ::  |:  ||   |    :|| : | ::
Db     2151 NNNN------QWLQIDLLRIKIKTTAITTQGCKSLSSEMYVKRYTIQYSDRGVENKSYREKSS 2206

Qy      411 NEEKVFQGNSNFRDPVQNNFIPPIVARVRVRVPQTWHQRIALKVELTGCOI 461
```

Db 2207 MVDKIFPEGNNRIGHVKHFNPPIISRFITRIPKWMNQSIALRLELFGCDI 2257

```
RESULT 12
ID   MFGM_HUMAN   STANDARD;   PRT;   387 AA.
AC   Q08431;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule factor 8) (MFG-E8) (HMFG)
DE   (Breast epithelial antigen BA46) (MFGM) (Contains: Medin).
GN   MFGE8.
OS   Homo sapiens (human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast, and Breast carcinoma;
RX   MEDLINE=96213908; PubMed=8639264;
RA   Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
RT   "Cloning and sequence analysis of human breast epithelial antigen
RT   BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT   growth factor-like domain.";
RL   DNA Cell Biol. 15:281-286(1996).
RN   [2]
RP   SEQUENCE OF 170-387 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=91371351; PubMed=1909932;
RA   Larocca D., Peterson J.A., Urrea R., Kuniyoshi J., Bistrain A.M.,
RA   Ceriani R.L.;
RT   "A Mr 46,000 human milk fat globule protein that is highly expressed
RT   in human breast tumors contains factor VIII-like domains.";
RL   Cancer Res. 51:4994-4998(1991).
RN   [3]
RP   PARTIAL SEQUENCE, AND CHARACTERIZATION.
RC   TISSUE=Milk;
RX   MEDLINE=98194924; PubMed=9535276;
RA   Giuffrida M.G., Cavaletto M., Giunta C., Conti A.,
RA   Godovac-Zimmermann J.;
RT   "Isolation and characterization of full and truncated forms of human
RT   breast carcinoma protein BA46 from human milk fat globule membranes.";
RL   J. Protein Chem. 17:143-148(1998).
RN   [4]
RP   SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
RX   MEDLINE=99342076; PubMed=10411933;
RA   Haeggvist B., Naeslund J., Sletten K., Westermark G.T., Mucchiano G.,
RA   Tjernberg L.O., Nordstedt C., Engstroem U., Westermark P.;
RT   "Medin: an integral fragment of aortic smooth muscle cell-produced
RT   lactadherin forms the most common human amyloid.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=97405885; PubMed=9260929;
RA   Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
RT   "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT   expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
RT   (RGD)-dependent cell adhesion.";
RL   DNA Cell Biol. 16:861-869(1997).
-!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
-!-   SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
-!- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
-!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
-!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
-!-   MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
-!- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
-!-   AMINO ACID 264 AND 273.
-!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
-!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
```

```
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U58516; AAC50549.1; -.
DR   EMBL; S65151; AAB19771.1; -.
DR   MIM; 602281; -.
DR   InterPro; IPR000561; EGF.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01285; FA58C_1; 1.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
FT   SIGNAL        1     23
FT   CHAIN        24    387       LACTADHERIN.
FT   CHAIN       202    387       LACTADHERIN; SHORT FORM.
FT   CHAIN       268    317       MEDIN.
FT   DOMAIN       24     67       EGF-LIKE.
FT   DOMAIN       70    225       F5/8 TYPE C 1.
FT   DOMAIN      230    387       F5/8 TYPE C 2.
FT   SITE         46     48       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     27     38       BY SIMILARITY.
FT   DISULFID     32     55       BY SIMILARITY.
FT   DISULFID     57     66       BY SIMILARITY.
FT   DISULFID     70    225       BY SIMILARITY.
FT   DISULFID    212    216       BY SIMILARITY.
FT   DISULFID    230    387       BY SIMILARITY.
FT   CARBOHYD    238            N-LINKED (GLCNAC).
FT   CARBOHYD    325            N-LINKED (GLCNAC).   (POTENTIAL).
FT   CARBOHYD    329            N-LINKED (GLCNAC).   (POTENTIAL).
FT   CARBOHYD    350            N-LINKED (GLCNAC).   (POTENTIAL).
SQ   SEQUENCE   387 AA;  43123 MW;  2EE657D3BCB37B2D CRC64;
```

```
Query Match             9.8%;   Score 299;   DB 1;   Length 387;
Best Local Similarity   37.3%;
Matches   63;   Conservative   33;   Mismatches   55;   Indels   18;   Gaps   5;

Qy  300  NGCSRSLSFE----PGGQIRASSWQSVNESGDQVH---WSPGQARLQDGG--PSWASGD  350
         ::  : :: :: :::   :: :::     :  :::::::   :::   :: :::
Db   228  NGCANPLGIKNNSIPDKQITASSSVKIWG----LHLFSNNPSYARLDKQGNFNAWVAGS  282

Qy  351  SSNHKPREWLEIDLGKEKKIIGIRITGSTQSNEFYVKSFEYMNFRKNNSKWRTYKGTVN  410
         ::::  ::  ::::::::::::  : ::::  :: : :::::::::  : :: :::: :
Db   283  YGND-----OWLQVDLGSSKEVTGIITQGARNFGSVQFVASYKVAYSNDSANWTEYQDPRT  338

Qy  411  NEEKVFQGNSNFRDPVQVNNFIPIVARYVRVPGTWHQRIALKVELIGC  459
         : :::: :::::::::::: :: ::  :: : :: ::: :::: :::
Db   339  GSSKIFPGNWDNSHSKKNLFETPLLARYVRIIPVAWHNRIALRLELLGC  387
```

```
RESULT 13
ID   MFGM_RAT   STANDARD;   PRT;   427 AA.
AC   P70490;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
DE   acetyl CD3 ganglioside synthase) (AGS) (MFGM).
GN   MFGE8 OR AGS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
```

Left column

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96374422; PubMed=8780713;
RA   Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
RL   "Cloning and expression of cDNA for O-acetylation of GD3
RL   ganglioside.";
RL   Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
CC       PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
```

```
DR   EMBL; D84068; BAA12210.1; -.
DR   HSSP; P00740; 1IXA.
DR   InterPro; IPR000561; EGF.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS00186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    427       LACTADHERIN.
FT   DOMAIN       24     61       EGF-LIKE 1.
FT   DOMAIN       64    108       EGF-LIKE 2.
FT   DOMAIN      111    267       F5/8 TYPE C 1.
FT   DOMAIN      272    427       F5/8 TYPE C 2.
FT   DISULFID     28     39       BY SIMILARITY.
FT   DISULFID     33     49       BY SIMILARITY.
FT   DISULFID     51     60       BY SIMILARITY.
FT   DISULFID     68     79       BY SIMILARITY.
FT   DISULFID     73     96       BY SIMILARITY.
FT   DISULFID     98    107       BY SIMILARITY.
FT   DISULFID    111    258       BY SIMILARITY.
FT   DISULFID    254    427       BY SIMILARITY.
FT   DISULFID    272    427       BY SIMILARITY.
FT   SITE         87     88       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD     51     61       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     61     61       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    230    230       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    390    390       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   427 AA;  47413 MW;  EA8C8631F3EE6047 CRC64;
```

```
Query Match  91; Similarity  26.0%; Score 297.5; DB 1; Length 427;
Best Local Similarity  44; Conservative  9.7%; Pred. No. 3.1e-15;
Matches  91; Mismatches  96; Indels 119; Gaps 12;

Qy  215 TEYSKFGPA-------------GCRDVAGDISGNMVDGYRDTSLLCKAAIHAGIIA-DELGSGOI 264
            |||                   |:|       ||  |:||| |:   |||||| |  ::||: |
Db   92 TEYICQCPVGYSGIHCELGCSTKLGLEGGAIAD-------SOISASSVYMGFMGLGRMGPEL 146

Qy  265 SVYLGRGI------------------SKYEGILANGVLSRDG-----------------SL 290
            ||:|||| 
Db  147 ... 205
```

Right column

```
RESULT 14

FA8_HUMAN
ID   FA8_HUMAN    STANDARD;    PRT;   2351 AA.
AC   P00451;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Coagulation factor VIII precursor (Procoagulant component)
DE   (Antihemophilic factor) (AHF).
GN   F8 OR F8C.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86081164; PubMed=3935400;
RA   Truett M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,
RA   Hartog K., Kuo C.H., Masiarz F.R., Merryweather J.P., Najarian R.,
RA   Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,
RA   Urdea M.S., Valenzuela P., Dahl H.-H.M., Pavalara J., Hansen J.,
RA   Nordfang O., Ezban M.;
RT   "Characterization of the polypeptide composition of human factor
RT   VIII:C and the nucleotide sequence and expression of the human
RT   cDNA.";
RL   DNA 4:333-349(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85061548; PubMed=6438526;
RA   Wood W.I., Capon D.J., Simonsen C.C., Eaton D.L., Gitschier J.,
RA   Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,
RA   Delwart E., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;
RT   "Expression of active human factor VIII from recombinant DNA clones.";
RL   Nature 312:330-337(1984).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85061550; PubMed=6438528;
RA   Toole J.J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,
RA   Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,
RA   Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,
RA   Hewick R.M.;
RT   "Molecular cloning of a cDNA encoding human antihaemophilic factor.";
RL   Nature 312:342-347(1984).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93265012; PubMed=1303178;
RA   Gitschier J., Wood W.I., ...
RT   "Sequence of the exon-containing regions of the human factor VIII
RT   gene.";
RL   Hum. Mol. Genet. 1:199-200(1992).
RN   [5]
RP   SEQUENCE OF 2064-2070 FROM N.A.
RA   de Water N.S., Williams R., Browett P.J.;
RL   Submitted (JUN-1997) to the EMBL/Genbank/DDBJ databases.
RN   [6]
RP   SULFATION OF TYR-1699.
```

```
Qy  291 SDKREFLFTS--------------------------------------------- 299
            ||:|:||:|
Db  206 DGRREFITQDSESGTGCKEFMCGNODDNNSLKINMENFPTLEAQYIRLYFVSCHRGCTLRFELL 265

Qy  300 ----NGCSRSLSFE----PDGGIRASSSMQSSWNESGDQVIHWSPGQARLQDQGP-SWASG 349
            |||  ||:| |  |||  ::||| ||| ||
Db  266 GCELHGCSEFGLGKNNTIPDSQITASSSYKTWNLRA--FGWYPHLGRLDNQGKINAW--- 320

Qy  350 DSSNNHKPRFWLEIDLGERKKITGIRTTGSTOSNFNFYVKSFYMNRFKNNNSKWKTYKGIV 409
            ||:|||:|||   | :::|:  |: :| | ::||| : ||| |::| : ||||:|| ||
Db  321 -TAQSNSAKEWLQVDLGTKKKVTGIITQGARDFEGHICTVASYKVAHSDDGVOWMTVTE--E 377

Qy  410 NNEEKVFQGNSNFRDPVQNFFIPPIVARYVRVPQTWHORIALAKVELLGC 459
            |||| | ::| |::|:| |||    :| |  | :||||:|| :|: |:||:|| |
Db  378 OGTSKVFQGNLDNHSHKKNIFEKPFMARYVRVPLPSWHNRITLRLELLGC 427
```

RX   MEDLINE-91093266; PubMed-1898735;
RA   Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
RA   Mertens K., van Mourik J.A.;
RT   "Sulfation of Tyr1680 of human blood coagulation factor VIII is
RT   essential for the interaction of factor VIII with von Willebrand
RT   factor.";
RL   J. Biol. Chem. 266:740-746(1991).
RN   [7]
RP   SULFATION.
RX   MEDLINE-92207952; PubMed-1554716;
RA   Pittman D.D., Wang J.H., Kaufman R.J.;
RT   "Identification and functional importance of tyrosine sulfate
RT   residues within recombinant factor VIII.";
RL   Biochemistry 31:3315-3325(1992).
RN   [8]
RP   STRUCTURE BY NMR OF 2322-2343.
RX   MEDLINE-95200924; PubMed-7893714;
RA   Gilbert G.E. Baleja J.D.;
RT   "Membrane-binding peptide from the C2 domain of factor VIII forms an
RT   amphipathic structure as determined by NMR spectroscopy.";
RL   Biochemistry 34:3022-3031(1995).
RN   [9]
RP   REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX   MEDLINE-91221499; PubMed-1902642;
RA   Gitschier J.;
RT   "The molecular basis of hemophilia A.";
RL   Ann. N.Y. Acad. Sci. 614:89-96(1991).
RN   [10]
RP   REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX   MEDLINE-89088506; PubMed-2491949;
RA   White G.C. II, Shoemaker C.B.;
RT   "Factor VIII gene and hemophilia A.";
RL   Blood 73:1-12(1989).
RN   [11]
RP   REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX   MEDLINE-96235434; PubMed-3012775;
RA   Gitschier J., Wood W.I., Shuman M.A., Lawn R.M.;
RT   "Identification of a missense mutation in hemophilia A.";
RL   Science 232:1415-1416(1986).
RN   [12]
RA   Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
RT   "Molecular etiology of factor VIII deficiency in hemophilia A.";
RL   Hum. Mutat. 5:1-22(1995).
RN   [13]
RP   VARIANT PRO-2135.
RX   MEDLINE-88096539; PubMed-3121181;
RA   Levinson B., Janco R.L., Phillips J.A. III, Gitschier J.;
RT   "A novel missense mutation in the factor VIII gene identified by
RT   analysis of amplified hemophilia DNA sequences.";
RL   Nucleic Acids Res. 15:9797-9805(1987).
RN   [14]
RP   VARIANT GLN-2228.
RX   MEDLINE-88191889; PubMed-2833855;
RA   Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
RA   Kazazian H.H.;
RT   "Nonsense and missense mutations in hemophilia A: estimate of the
RT   relative mutation rate at CG dinucleotides.";
RL   Am. J. Hum. Genet. 42:718-725(1988).
RN   [15]
RP   VARIANT GLY-291.
RX   MEDLINE-88220354; PubMed-2835904;
RA   Youssoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr.,
RA   Antonarakis S.E.;
RT   "Moderately severe hemophilia A resulting from Glu-->Gly substitution
RT   in exon 7 of the factor VIII gene.";
RL   Am. J. Hum. Genet. 42:867-871(1988).
RN   [16]
RP   VARIANT CYS-1708.
RA   O'Brien D.P., Tuddenham E.G.;

RT   "Purification and characterization of factor VIII 1,689-Cys: a
RT   nonfunctional cofactor occurring in a patient with severe hemophilia
RT   A.";
RL   Blood 73:2117-2122(1989).
RN   [17]
RP   VARIANT CYS-391.
RA   Shima M., Ware J., Yoshioka A., Fukui H., Fulcher C.A.;
RT   "An arginine to cysteine amino acid substitution at a critical
RT   thrombin cleavage site in a dysfunctional factor VIII molecule.";
RL   Blood 74:1612-1617(1989).
RN   [18]
RP   VARIANT LEU-189.
RX   MEDLINE-90057680; PubMed-2510835;
RA   Chan V., Chan T.K., Tong T.M., Todd D.;
RT   "A novel missense mutation in exon 4 of the factor VIII gene
RT   resulting in moderately severe hemophilia A.";
RL   Blood 74:2688-2691(1989).
RN   [19]
RP   VARIANT HIS-391.
RX   MEDLINE-89197216; PubMed-2495245;
RA   Inaba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
RT   "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26
RT   of the factor VIII gene.";
RL   Hum. Genet. 81:335-338(1989).
RN   [20]
RP   VARIANT LEU-2326.
RX   MEDLINE-89264602; PubMed-2498882;
RA   Arai M., Inaba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
RA   Fujimaki M., Hoyer L.W.;
RT   "Direct characterization of factor VIII in plasma: detection of a
RT   mutation altering a thrombin cleavage site
RT   (arginine-372-->histidine).";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
RN   [21]
RP   VARIANT CYS-1708.
RX   MEDLINE-90123183; PubMed-2105106;
RA   Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
RA   Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
RT   "Recurrent mutations and three novel rearrangements in the factor
RT   VIII gene of hemophilia A patients of Italian descent.";
RL   Br. J. Haematol. 75:73-77(1990).
RN   [23]
RP   VARIANT CYS-391.
RX   MEDLINE-90329422; PubMed-1973901;
RA   Pattinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
RT   "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the
RT   internal heavy chain thrombin cleavage site.";
RL   Blood 75:384-389(1990).
RN   [24]
RP   VARIANTS GLN-2228 AND LEU-2326.
RX   MEDLINE-90152691; PubMed-2105906;
RA   Higuchi M., Wong C., Kochhan L., Olek K., Aronis S., Kasper C.K.,
RA   Kazazian H.H., Antonarakis S.E.;
RT   "Characterization of mutations in the factor VIII gene by direct
RT   sequencing of amplified genomic DNA.";
RL   Genomics 6:65-71(1990).
RN   [25]
RP   VARIANTS PHE-1699 AND CYS-1708.
RX   MEDLINE-91069988; PubMed-2106480;
RA   Kazazian H.H., Higuchi M., Kasper C.K., Antonarakis S.E.,
RA   Traystman M.D.;
RT   "Use of denaturing gradient gel electrophoresis to detect point
RT   mutations in the factor VIII gene.";

Left column

```
Query Match           9.4%;  Score 285.5;  DB 1;  Length 2351;
Best Local Similarity    37.6%;  Pred. No. 2.9e-21;
Matches   62;  Conservative   32;  Mismatches   56;  Indels   13;  Gaps   4;

Qy   300  NGCGSRSLSFE-----PDQGIRASSWSMQSVNESGDQVHNSPGQARLQDQGPSWASGDSSNNH 355
               |||||| |                    | |||:|    |        ||    ::::||
Db  2191  NSCSMPLGMESKAISDAQITASSYFTNMFAT-----WSPSKARLHLQGRSNAMRPQVNN- 2244

Qy   356  KPREWLEIDLGEKKKITGIRTTGSTQDSNFNFYVKSFVVMNFKNNNSKWKTYKGIVNNEEKY 415
                |:|||:|          | |  |         | | :      |::|:|:| |
Db  2245  -PKEWLQVDFQGKTMKYTGVTTQGVKSLLTSMYVKEFLISSSQDSGHQWTLF--FQNGKVKY 2301

Qy   416  FQGNSNFRDPVQNNFIPPIVARYVRVPQTWHQRIALKVELIGCQ 460
               ||||:::|| |||||:|| |||||:|:||:|||||||||||||||||
Db  2302  FQGNDDSFTPVVNSLDPLLITRYLRIHPQSWVHQIALRMEVLGCE 2346
```

Right column

```
Query Match           9.1%;  Score 276.5;  DB 1;  Length 463;
Best Local Similarity    23.4%;  Pred. No. 1.4e-13;
Matches  112;  Conservative   59;  Mismatches  118;  Indels  189;  Gaps   21;

Qy   141  CASDYLFTSSSDQYCGYCGSMTVPKRLLLNTSEVTVRFESGSHISGR-------GF 190
                 |  |:|   |::|          |:|      |  | |||||||:|:|:|         ||
Db    16  CASG---LFPASGD-----FCDS------SLCLN--------GGTCLTGQDNDIYCLCPEGF 55

Qy   191  LLIYAS--------------SDHPDLITCLERASHYLKTEYSKFCPAGCRDVAGDISGNM- 236
                 ||:        |  |: |  | |::||   |  | |:  ::| |   | |||:|:|:|
Db    56  TGLVCNETERGCPSPNPCYNDAKCLVT-LDTPQRGDIFTEYICQCPGYSGIHCETETNY 114

Qy   237  ---VDG-YRDTSLLCKAAI------------------------HAGIIAD----------- 258
                  | |            |                              |:|:|
Db   115  NLDGEYMFTIVAPVNTNAVPTPAPTPDLSNNIASRCGSTQLGMEGGAIASDSQISASYYWGFN 174

Qy   259  ---ELGGQISVLQRKGI--------------------------SRYEGILANGVLSRDG----- 288
                  ||  | :| ||| |                            || |:| |: | |:|
Db   175  GLDRMGQPELARLFRTGIVNAHHASNTDSKPWIQVNLLRKMRVSGVMTQGA-SRAGRAEYL 233

Qy   289  -----SLSDKRFLFTSN-------------------------------GCSR- 304
                    ||:|:| |||:|                                  ||::|
Db   234  KTFKVAYSLDGKRFEFIQDESGGDKEFLGNLDNSLKVNMFNFTLEAQIIRLYPVSCHRG 293

Qy   305  --SLSFE-----------------------------PDGQIRASSWSVNESGDQVHNSPGQARLQDQG 343
                  ||:||                             ||||||||||||||||||||||||||||||||
Db   294  CTLRFELLGCELHGCLEPLGLKNNTIPDSQMSASSYKTWNLRA--FGWPHLGRLDNQG 351

Qy   344  P--SWASGDSSNHHIKPREWLEIDLGEKKKITGIRTTGSTQDSNFNFYVKSFVANFKNNNSK 401
                   :  |||||||::|||||||||||||||||||||:||||| :|:|||:|| |:||| ||||
Db   352  KINAW-----TAQSNSAKEWLQVDLGTRQVQVGIITQGARDFGHIIQYVESKVKVAHSDGQVQ 407
```

Lower left column (SWISS-PROT entry)

```
RESULT 15

MFGM_MOUSE

ID   MFGM_MOUSE     STANDARD;     PRT;     463 AA.
AC   P21956;  P97800;
DT   01-AUG-1991  (Rel. 19, Created)
DT   01-AUG-1991  (Rel. 19, Last sequence update)
DT   16-OCT-2001  (Rel. 40, Last annotation update)
DE   Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
DE   (Sperm surface protein SP47) (MP47).
GN   MFGE8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. AND SEQUENCE OF 23-35.
RC   TISSUE=Mammary gland;
RA   Ensslin M.A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 23-463 FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=91046008; PubMed=2122462;
RA   Stubbs J.D., Lekutis C., Singer K.L., Bui A., Yuzuki D.,
RA   Srinivasan U., Parry G.;
RT   "cDNA cloning of a mouse mammary epithelial cell surface protein
RT   reveals the existence of epidermal growth factor-like domains linked
RT   to factor VIII-like sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
CC   -!- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC       BINDING PROTEIN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC       SPERMATOZOAN.
CC   -!- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
CC       TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC   -------------------------------------------------------------------
```

```
DR   EMBL; M38337; AAA39534.1; -.
DR   EMBL; Y11684; CAA72380.1; -.
DR   PIR; A36479; A36479.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:1022768; Mfge8.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00754; F5_F8_type_C; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01285; FA58C_1; 2.
DR   PROSITE; PS01286; FA58C_2; 2.
KW   Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
FT   SIGNAL        1     22
FT   CHAIN        23    463       LACTADHERIN.
FT   DOMAIN       24    108       EGF-LIKE 1.
FT   DOMAIN       64    108       EGF-LIKE 2.
FT   DOMAIN      148    303       F5/8 TYPE C 1.
FT   DOMAIN      303    463       F5/8 TYPE C 2.
FT   SITE         87     89       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     28     39       BY SIMILARITY.
FT   DISULFID     33     49       BY SIMILARITY.
FT   DISULFID     51     60       BY SIMILARITY.
FT   DISULFID     68     73       BY SIMILARITY.
FT   DISULFID     73     96       BY SIMILARITY.
FT   DISULFID     98    107       BY SIMILARITY.
FT   DISULFID    148    303       BY SIMILARITY.
FT   DISULFID    290    294       BY SIMILARITY.
FT   DISULFID    308    463       BY SIMILARITY.
FT   CARBOHYD     61     61       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    266    266       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    426    426       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     30     30       S -> F (IN REF. 2).
FT   CONFLICT     35     35       N -> D (IN REF. 2; AA SEQUENCE).
FT                                EINYYINLDSEYMFTIAVPNTAVPETPAPTPDLSNNLIASR ->
FT   CONFLICT    110    147       G (IN REF. 2).
FT   CONFLICT    168    168       Y -> S (IN REF. 2).
FT   CONFLICT    196    196       H -> Y (IN REF. 2).
FT   CONFLICT    309    309       L -> S (IN REF. 2).
FT   CONFLICT    309    309       E -> A (IN REF. 2).
SQ   SEQUENCE   463 AA;  51465 MW;  D7B8GC6EFBBA724D CRC64;
```

```
Qy  402 WKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVPQTWHQRIALKVELIGC 459
        |:|   ||||||:  ::||:| |  |:||||||||:| :|| || |::||:||:||
Db  408 WTVYE--EQGSSKVFQGNLDNNSHKNIFEKPFMARYVRVLPVSWHNRITLRLELLGC 463
```

Search completed: July 9, 2002, 02:26:11
Job time: 648 sec

OM protein - protein search, using sw model

Run on: July 9, 2002, 02:14:03 ; Search time 130.73 seconds
                 [without alignments]
                 775.433 Million cell updates/sec

Title: US-09-691-344a-4
Perfect score: 3052
Sequence: 1 MGFGAGQRLRPVPAPRSSAE..................YDNEKEMTQKLDLTSDMAG 586

Scoring table: BLOSUM62
                 Gapop 10.0 ; Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_19:*
                 1: sp_archea:*
                 2: sp_bacteria:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                 10: sp_plant:*
                 11: sp_rodent:*
                 12: sp_virus:*
                 13: sp_vertebrate:*
                 14: sp_unclassified:*
                 15: sp_rvirus:*
                 16: sp_bacteriap:*
                 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                                  GenCore version 4.5
                              Copyright (c) 1993 - 2000 Compugen Ltd.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1310.5	42.9	503	11	Q9D4J3	mus musculu
2	1154.5	37.8	460	11	Q9D696	mus musculi
3	942	37.0	460	11	Q9I1V2	rattus norv
4	937	30.7	769	11	Q9I1V3	mus musculu
5	929	30.7	769	4	Q96PD2	homo sapien
6	572.5	18.8	364	4	O14089	homo sapien
7	349	11.4	555	4	Q9H2E2	homo sapien
8	349	11.4	901	4	Q9H2E4	homo sapien
9	349	11.4	901	4	Q9H2E3	homo sapien
10	349	11.4	906	4	Q9H2D5	homo sapien
11	349	11.4	906	4	Q9H2D4	homo sapien
12	335	11.0	609	4	Q96I90	homo sapien
13	335	11.0	644	4	Q96IH5	homo sapien
14	335	10.7	704	4	Q96I90	homo sapien
15	325.5	10.5	921	11	Q90X38	rattus norv
16	321	10.5	2224	4	O43737	homo sapien

17	318	10.4	2183	11	O87783	mus musculu
18	305.5	10.0	2343	6	O18806	canis famil
19	305.5	10.0	2343	6	O62730	canis famil
20	299.5	9.8	779	4	Q9NTt3	homo sapien
21	290	9.4	426	11	Q9WTS3	mus musculu
22	285.5	9.4	216	4	O14286	homo sapien
23	279.5	9.2	463	13	Q9R1X9	homo sapien
24	277	9.1	2119	13	Q90X47	brachydanio
25	268	8.8	480	11	O43854	mus musculu
26	268	8.7	480	11	O35474	mus musculu
27	238.5	7.8	363	11	O77718	mus musculu
28	226	7.4	1128	4	O88442	homo sapien
29	221.5	7.3	224	11	O88442	mus musculu
30	217	7.1	1158	4	O91RM6	mus musculu
31	216.5	7.1	719	4	O14113	homo sapien
32	209.5	6.9	858	5	O61281	drosophila
33	199.5	6.5	728	9	O97567	homo sapien
34	197	6.5	721	4	O97470	lytechinus
35	194.5	6.4	335	4	O14113	homo sapien
36	188.5	6.2	3843	5	O96M80	drosophila
37	188.5	6.2	3843	5	O9BTL9	drosophila
38	186.5	6.1	745	4	O9USD0	homo sapien
39	186.5	6.1	963	5	O9VV94	drosophila
40	185.5	6.1	764	4	O96MJ5	homo sapien
41	182.5	6.0	764	11	O96MS7	mus musculu
42	182.5	6.0	604	4	O96215	homo sapien
43	182	6.0	764	11	O54860	mus musculu
44	173.5	5.7	1524	13	O91674	xenopus lae
45	173	5.7	926	11	Q9U000	homo sapien
					Q9uQ0	homo sapien
					Q9cx06	mus musculu

ALIGNMENTS

RESULT 1
Q9D4J3
ID   Q9D4J3;   PRELIMINARY;   PRT;   503 AA.
AC   Q9D4J3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   4631413K11RIK PROTEIN.
GN   4631413K11RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085560; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   MEDLINE=21085560; PubMed=11217851;
RA   Kadota H., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RL   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR   EMBL; AK016485; BAB30265.1; -.

```
DR    MGI; MGI:1913336; 4631413K11Rik.
DR    InterPro; IPR000859; CUB.
DR    InterPro; IPR004043; LCCL.
SMART; SM00042; CUB; 1.
Pfam; PF00431; CUB; 1.
PROSITE; PS01180; CUB; 1.
SQ    SEQUENCE 503 AA;  54547 MW;  FEB121E845CAA06B CRC64;

Query Match     42.9%;  Score 1310.5;  DB 11;  Length 503;
Best Local Similarity   51.0%;  Pred. No. 2.5e-96;
Matches  268;  Conservative  26;  Mismatches  32;  Indels   1;

Qy     61  AAGRGLLALILAVSAPLRLQAEELGDGCGHLVTYQDSGGTMVTSKNYPGTYVNHYVCEKIIT 120
               :|:::||  |||||       | |:|| |:   |||| :|:||||||||||||||||||
Db      5  AGGPSVIALLFAVCAAPLRLQAEELGDGCGHIVTSQDSGGTMTSKNYPGTYVNHYVCEKIT 64

Qy    121  VPKGKRLILRLGDLDIESQTCASDYLLFTSSSDQYGPYCCSSMTVPKRELLLNTEEKTVRFE 180
               ||||||||||||||||||:||||||||||::||:|||||||||  |  |:|::||: |
Db     65  VPKGKRLILRLGDLNIESKTCASDYLLFSSATDQYGPYCCSSWAVPKELRLNSNEVVTVLFK 124

Qy    181  SGSHISGRGFLITYASSDHPDLITCLERASHYLKTEIYSKRFCPAGCRDVAGDIGSGNMVDGI 240
               |||||||||||||||||:||||||||||||:|:|:|::| :||||||||||||||::||||
Db    125  SGSHISGRGFLITYASSDHPDLITCLERGSHYFEEKYSKFCPAGCRDIAGSGNTKDGY 184

Qy    241  RDTSLLCKRAAITHAGITIADELGGQISVLQRKGISRYEGILANGVLSRDGSLSDKRFLFTSN 300
               |||||||||||:||:||:||||:|||::||||::|||||||:|||||::|||||||:|||:|
Db    185  RDTSLLCKRAAIHAGIITDELGGHINLLQSKGISHYEGLLANGVLSRHGSLSEKRFLP--- 241

Qy    301  GCSRSLSFEEPDQGIRASSSWQSVNESGDQVHWSPGQARLQDQGPSWASGCDSSNNHKPREW 360
                                                                            241

Qy    361  LEIDLGEKKKIGIRTGTGSTQSNFNFVYKSFVWNNFKNNNSKWKTYKGIVNNEEKVFQGNS  420
                                                                            241

Qy    421  NFRDPVQNNFIPPIVARYVRVPQTWHQRIALKVELIGCCITOGNDSLVWRKTSQSTSVS  480
                                                                            241

Qy    481  TKKEDETIIRPIPSEETSTGINITTVAIPLVLLVLVFPAGMGIFAAFRKKKKGSPYGSA  540
                                                                            241

Qy    541  EAQKTDCWKQIKYPFARHQSAEFFISYDNEKEMTQKLDLITSDMA  585
                                                                            241

Db    242  --------------------------------------------  241

Db    282  DAQKTSCWKQIKYPFARHQSTEFFISYDNEKEMTQKLDLITSDMA  330
```

```
RESULT   2
ID    Q90696
AC    Q90696;
DT    01-JUN-2001 (TrEMBLrel. 17, Created)
DT    01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT    01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE    4631413K11RIK PROTEIN.
GN    4631413K11RIK.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX    NCBI_TaxID=10090;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=C57BL/6J; TISSUE=SKIN;
RX    MEDLINE=21085660; PubMed=11217851;
RA    Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA    Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA    Aizawa M., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA    Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA    Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA    Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
```

```
RA    Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA    Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA    Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA    Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA    Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA    Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA    Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA    Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA    Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA    Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA    Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA    Hayashizaki Y.;
RT    "Functional annotation of a full-length mouse cDNA collection.";
RL    Nature 409:685-690(2001).
CC    -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR    EMBL; AK014521; BAB29409.1; -.
DR    MGD; MGI:1913336; 4631413K11Rik.
DR    InterPro; IPR000859; CUB.
DR    InterPro; IPR004043; LCCL.
SMART; SM00042; CUB; 1.
Pfam; PF00431; CUB; 1.
PROSITE; PS01180; CUB; 1.
SQ    SEQUENCE 460 AA;  50334 MW;  3CF3356F70A88B93 CRC64;

Query Match     37.8%;  Score 1154.5;  DB 11;  Length 460;
Best Local Similarity   48.8%;  Pred. No. 6.5e-84;
Matches  237;  Conservative  24;  Mismatches  26;  Indels 199;  Gaps   1;

Qy    100  MTSKNYPGTYVNHYVCEKIITVPKGKRLILRLGDLDIESQTCASDYLLFTSSSDQYGPYC 159
               ||||||||||||||||||||||||||||||||||||:||||||||||::||:||||||
Db      1  MTSKNYPGTYVNHYVCEKITVPKGKRLILRLGDLNIESKTCASDYLLFSSATDQYGPYC 60

Qy    160  GSMTVPKRELLLNTSEVTVRFESGSHISGRGFLITYASSDHPDLITCLERASHYLKTEYSK 219
                |:  |:|::||: |    |||||||||||||||||||:||||||||||:|:|:|::|
Db     61  GSWAVPKELRLNSNEVVTVLFKSGSHISGRGFLITYASSDHPDLITCLERGSHYFEEKYSK 120

Qy    220  FCPAGCRDVAGDIGSGNMVDGIRDTSLLCKRAAIHAGIIADELGGQISVLQRKGISRYEGIL 279
               ||||||||:||:|::||||||||||||||||||:||:||||:||::||||::|||||||
Db    121  FCPAGCRDIAGSGNTKDGYRDTSLLCKRAAIHAGIITDELGGHINLLQSKGISHYEGLL 180

Qy    280  ANGVLSRDGSLSDKRFLFTSNGCSRSLSFEEPDQGIRASSSWQSVNESGDQVHWSPGQARL 339
               |||||||:|||||:|||:|                                          
Db    181  ANGVLSRHGSLSEKRFLP-- 198

Qy    340  QDQGPSWASGCDSSNNHKPREWLEIDLGEKKKIGIRTGTGSTQSNFNFVYKSFVWNFKNNN 399
                                                                            198

Qy    400  SKWKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVPQTWHQRIALKVELIGC 459
                                                                            198

Qy    460  QITTOGNDSLVWRKTSQSTSVSTIKEDEIIIRPIPSEETSTGINITTVAIPLVLLVLVFA 519
                                                                            198

Qy    520  GMGIFAAFRKKKKKGSPYGSAEAQKTDCWKQIKYPFARHQSAEFFISYDNEKEMTQKLDL 579
                                                                            198

Db    199  --------------------------------------------------------  198

Db    222  GMGIFAICKRRKKKGNPYVSADAQKTGCWKQIKYPFARHQSTEFFISYDNEKEMTQKLDL 281
```

```
RESULT   3
ID    Q91EV2
AC    Q91EV2;
DT    01-DEC-2001 (TrEMBLrel. 19, Created)
DT    01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT    01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

PRELIMINARY;  PRT;  769 AA.

Qy    580  ITSDMA  585
               ||||||
Db    282  ITSDMA  287
```

```
PRELIMINARY;  PRT;  460 AA.
```

RESULT 4

```
Q912V3
ID   Q912V3      PRELIMINARY;   PRT;   769 AA.

DE   ENDOTHELIAL AND SMOOTH MUSCLE CELL-DERIVED NEUROPILIN-LIKE
DE   PROTEIN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
GN   ESDN.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   PubMed=11447234;
RA   Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA   Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT   "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT   Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT   Up-regulated after Vascular Injury.";
RL   J. Biol. Chem. 276:34105-34114(2001).
DR   EMBL; AF387549; AAL30180.1; -.
SQ   SEQUENCE 769 AA; 83915 MW; 5B79396C8FA063A93 CRC64;

Query Match             30.9%;  Score 942;  DB 11;  Length 769;
Best Local Similarity   37.5%;  Pred. No. 1.4e-66;
Matches  230;  Conservative   96;  Mismatches  204;  Indels   84;  Gaps   19;

Qy   44 PSGVWVPGARGGALARA------GRG--------LLAILLAVSAPLRLQAEELGDGCG   89
         |||   ||:: :                       :   | ||||||| || | |||
Db   16 PGGRAAPAATGRAPLPSAGWCPLPPGFRNSSSSRPRLILLLLLLPDA-----SQTCASDYL-L  70

Qy   90 HLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPGKRLILRLGDLDIE-SQTCASDYL-L  147
Db   71 HVFLGPESGTLTSINYPHTYPNSTVCKWEIRVATGERIRIKFGDFDIEDSDYCHLNTLKI  130

Qy  148 FTS----SSDQYGPYCG--SMTVPKELLLNTSEVTVRFESSGHISGRGFLLTYASSDHPDLI  203
Db  131 FNGIGVSRTEICKYCGLGLQMNOSIESKGSEITVLPMSGIHASGRGFLASYSVIDKODLI  190

Qy  204 TCLERASHYLKTEYSKFCPAGCRDVAGDIGNMVDGYRDTSLICKAAIHAGIIADELGGQ  263
Db  191 TCLDIVSNFLPEFSKYCPAGCLLPFAEISGTIPHGYRDSSPLCMAGIHAGVVSDVLGGQ  250

Qy  264 ISVLQRRGIGRYEGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE------PDGQIRASSS  319
Db  251 ISVVISKGTPYYESSLANNVTSMVGVLSTSLFTFKTSGCYGTLGMESSGVIADPOITASSV  310

Qy  320 MOSVNESGDQVHWSRPGQARLQDQGPSWASGDSSNHKPREWLEIDLGEKKKITGIRTTGS  379
Db  311 LEWTIDHMGQENSWPEPKARLRKPGPPWAA-FATDEH------OMLQIDLNKEKKITGIVTTGS  366

Qy  380 TOSNFNFYVKSFPVNNFKNNNSKWIYKGIIVNNEEKVFQGNSNFRDPVQNNFIPIVARYV  439
Db  367 TLIEHNYYVSAYRVLYSDDGQRWTVYREPGQAQDKIFQGNKDVHKDVRNNFLPPIIAREI  426

Qy  440 RVVPGTWHQRTALKVELIGCGIT----QGNDSLVWRKTSQSTSVSTKKED  485
Db  427 RVNFVQWOQKIAMKVELGCGQFLKGRLPKLTQPPPRNSNNL--KNTVHPKLGRAPKF  484

Qy  486 EFITRP-----IPSEETST-GINITT----VAIPLVLLVVILVFAGMGIFAF-  527
Db  485 TOALQPRSRNDLPLLPAQTTATPDVKWKFCVKQIK--YPFARHQSAEFTISYONE----------K  571

Qy  528 RKKKKKGSPYGSAEAOKTDCWKQIK----VAIPLVLLVVILVFAGMGIFAF-  571
Db  545 CAWHWRNRKKAKGT-YDLPHWDRAGWWRKGVKQLLPAKSVEHEETPVRYSNSMWSHLSPSR  603

Qy  572 EMTQKDLDITSDMA  585
Db  604 EVTTVLQADSAEYA  617
```

RESULT 4

```
Q912V3
AC   Q912V3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENDOTHELIAL AND SMOOTH MUSCLE CELL-DERIVED NEUROPILIN-LIKE
DE   PROTEIN.
GN   ESDN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RX   PubMed=11447234;
RA   Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA   Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT   "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT   Cells with the Longest Secretory Signal Sequence among Eukaryotes, Is
RT   Up-regulated after Vascular Injury.";
RL   J. Biol. Chem. 276:34105-34114(2001).
DR   EMBL; AF387548; AAL30179.1; -.
SQ   SEQUENCE 769 AA; 83374 MW; 73C1F1646FA3F017 CRC64;

Query Match             30.7%;  Score 937;  DB 11;  Length 769;
Best Local Similarity   37.2%;  Pred. No. 3.6e-66;
Matches  229;  Conservative  100;  Mismatches  200;  Indels   86;  Gaps   17;

Qy   44 PSGVWVPGARGGALARA------GRG--------LLAILLAVSAPLRLQAEELGDGCGCHL  91
         |||   ||:: :                       :   | ||||||| || | |||
Db   16 PGGPAAPAATGRAALPSAGCCPLPPGRNSSSSRPRLILLLLLL-----LQDAGQQGCDGCGHT  72

Qy   92 VTYQDSGTMTSKNYPGTYPENHTVCEKTITVPGKRLILRLGDLDIE-SQTCASDYL-LFT  149
Db   73 VLGPESGTLTSINYPHTYPNSTVCEWEIRVRGERIRIKFGDFDIEDSDYCHLNTLKIFN  132

Qy  150 S----SSDQYGPYCG--SMTVPKELLLNTSEVTVRFESSGHISGRGFLLTYASSDHPDLITC  205
Db  133 LERASHYLKTEYSKFCPAGCRDVAGDISGNMVDGYRQTSLLCKAAIHAGIIADELGGQLITC  192

Qy  206 LERASHYLKTEYSKFCPAGCRDVAGDISGNMVDGYRQTSLLCKAAIHAGIIADELGGQ  265
Db  193 LDTVSNFLPEFSKYCPAGCLLPFAEISGTIPHGYRDSSPLCMAGIHAGVVSNVLGGIS  252

Qy  266 VLQRKGIGRYEGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE------PDGQIRASSWQ  321
Db  253 IVISKGTPYYESSLANNVTSVGVLSASLFTFKTSGCYGTLGMESSGVIADPOITASSALE  312

Qy  322 SVNESGDQVHWSPCQQARLQDQGPSWASGDSSNHKPREWLEIDLGEKKKITGIRTTGSTQ  381
Db  313 WTDHMGQENSWPAEKARLRKPGPPWAA-FATDEH------OMLQIDLNKEKKITGIVTTGSTM  368

Qy  382 SNFNFYVKSFYMNEFKNNNSKKKTYKGIIVNNEEKVPQGNSNFRDPVQNNFIPIVARYV  441
Db  369 IEHNSYYVSAYRVLYSDDGQRWTVYREPGVDQDKIFQGNKDVHKDVRNNFLPPIIARFIRV  428

Qy  442 VPGTWHQRIALKVELIGCGIT----QGNDSLVWRKTS-  474
Db  429 NPVQWQQKIAMKVELGCGQFLKGRLPKLTQPPPRNGNN--LRNTVARPKLGKRAPKFT  485

Qy  475 -----QSTSVSTKKEDEFITRPIPSEETSTGINITTVAIPLVLVLVVFAGMGIFAF-  527
Db  486 QVLQPRSRNELPVQPAETTTFPDIKNTTVTPSVIKDVALAAVIVPVLVMALTTILIIVC  545

Qy  528 -----RKKKKKGSPYGSAEAOKTDCWKQIKYPF--------------ARHQSAEFTISYDNE  570
Db  546 AWHWRNRKKYTEGA-YDLPHWDRAGWWGMKQLLPAKSVDHEEPVRYSTSE--VSHLSA  602

Qy  571 KEMTQKDLDITSDMA  585
Db  603 REVTTVLQADSAEYA  617
```

RESULT 5

```
Q96PD2
ID   Q96PD2;      PRELIMINARY;   PRT;   775 AA.
AC   Q96PD2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ENDOTHELIAL AND SMOOTH MUSCLE CELL-DERIVED NEUROPILIN-LIKE
DE   PROTEIN.
GN   ESDN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=11447234;
RA   Kobuke K., Furukawa Y., Sugai M., Tanigaki K., Ohashi N.,
RA   Matsumori A., Sasayama S., Honjo T., Tashiro K.;
RT   "ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
RT   Cells with the Longest Secretory Signal Sequence among Eukaryotes. Is
RT   Up-regulated after Vascular Injury.";
RL   J. Biol. Chem. 276:34105-34114(2001).
DR   EMBL; AF387547; AAL30178.1; -.
SQ   SEQUENCE 775 AA; 85034 MW; 3D06F81EE2337010 CRC64;
```

Best Local Similarity   30.4%;   Score 929; DB 4; Length 775;
Matches 225; Conservative 106; Mismatches 205; Indels 84; Gaps 17;

```
Qy   36  EAEIAKLAPSGVVMPGARGGGALARAAGRGL---LAILLAVSAPIRLQAEELGDCGHLV  92
Db   17  QVRAAAAAAWAALPLSRSLPPCSNSSSFSMPLFLLLLLVLILLLEDAGAQQODGCGHTV  76

Qy   93  TYQDSGTMTSKNYPGTYPNHTVCEKTITVPKGKRLILRLGDLDIE-SQTCASDYLLFTS-  150
Db   77  LGPESGTLTSINYPQTYPNSTVCEMEIRVKAMGERVRIKFGDPEDIEDSDSCHFNYLRIYG  136

Qy  151  ---SSDQYCPYCG--SMTVPKELLLNTSEVTAVRPFESGSHISGRGFLLTYASSDHPDLICL  206
Db  137  IGYSRTEICKYCGLGLQMHISIESKGNEITILLFMSGIHVSGRGFLASYVIDKQDLITCL  196

Qy  207  ERASHLKTEYSKFCPAGCRDVAGDISGNMVDSYRDTSLLCKRAAIHAGIIADELGGQISV  266
Db  197  DTASNLEPEFSKYCCPAGCLLPFAEISGTIPHGYRDSSPLCMAGVHAGVVSNTLGGQISV  256

Qy  267  LQRGLSRYEGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE------PDGQIRASSSWQS  322
Db  257  VISKGIPYYESSLANNVTSVVGHLSTSLFTFKTSGCYTLGMES-----GVIADPQITASSVLEW  316

Qy  323  VNESGDQVHWSPGQARLQDQGPSWASGDSSNNHKPREWL  382
Db  317  TDHTGQENSWKPKARULKRKPGPPWAAFATDE----YQWL  372

Qy  383  NFNFYVKSFVMNFKNNNSKWKIYKGIVNNEEKVFQGINSNFRDPVQNNFIPPIVAARVVRVV  442
Db  373  EHNIYVSATFILIYSDDGQKWTVYREPGVEQDKIFQGNHDYHDQVRNNFLPPIIARPIRVN  432

Qy  443  PQTWHQRIALKVELIGCQ------------ITQ------GNDSLVRKTSGQSISVSTKKEDET  487
Db  433  PYQWQQKIAMKMELLGCQFIFKGRPPKLTQQPPPRNSND-----LKRWTTAPPKIKAGRAPK  488

Qy  488  ITRPI--------------PSEET--STGINITT-----YAIPLVLILVLVFPAGMGIFFAAF  527
Db  489  FTQPLQPRSSNEFPAQTEQTTASPDIRNTTVFTPNVTKDVALAAVLVPVLVMVILTILILL  548

Qy  528  ---------RKKKKKGS----PYGSAEAQKTDCWKQIKKPF----------ARHQSAEFTI  565
Db  549  VCAAHWRNKKKTEGTVDLPYW----DRAGWKRGHKQFLPAKAVDHEETPVRSSSSE--V  602

Qy  566  SYDNEKEMTQKLDLITSDMA  585
Db  603  NHLSPREVTIVLQADSAEVA  622
```

RESULT 6

```
Q14089
ID   Q14089;      PRELIMINARY;   PRT;   364 AA.
AC   Q14089;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shibata T.;
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RL   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
RT   "unpublished.";
DR   EMBL; D29810; BAA19909.1; -.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR004043; LCCL.
DR   Pfam; PF00431; CUB; 1.
DR   Pfam; PF00754; F5_F8_type_C; 1.
DR   SMART; SM00042; CUB; 1.
DR   PROSITE; PS01180; CUB; 1.
KW   Hypothetical protein.
FT   NON_TER   1
SQ   SEQUENCE 364 AA; 39954 MW; FECA6292E9E0801F CRC64;
```

Best Local Similarity   18.8%;   Score 572.5; DB 4; Length 364;
Matches 147; Conservative 61; Mismatches 124; Indels 49; Gaps 16;

```
Qy   86  DGCGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPKGKRLILRLGDLDIE-SQTCASD  144
Db    1  DGCGHTVLGPESGTLTSINYPQTYPNSTVCEMEIRVKAMGERVRIKFGDPEDIEDSDSCHFN  60

Qy  145  YLLFTS----SSDQYCPYCG--SMTVPKELLLNTSEVTAVRPFESGSHISGRGFLLTYASSDH  199
Db   61  YLRIYGIGYSRTEICKYCGLGLQMHISIESKGNEITILLFMSGIHVSGRGFLASYVID-  119

Qy  200  PDLICLERASHLKTEYSKF-----------CPAGCRDVAGDISGNMVDSYRDTSLLCKRAA  250
Db  120  -----KQRSNLFGHCIQFFWNLSSVSTAPAGCLLPFAEISGTIPHGYRDSSPLCMAG  172

Qy  251  IHAGIIADELGGQISVLQRGLSRYEGILANGVLSRDGSL-SDKRFLFTSNGCSRSLSFE  309
Db  173  VHAGVVSNTLGGQISVVISKGIPYYESSLANNVTSVVGHLILQVFFTFKTSGCYTLGME  232

Qy  310  PDGQIRASSS------WQSNSNEGSDDQVHWSPGQARLQDQGPSW-ASGDSSNNHKPREWL  361
Db  233  SGGD-RGSSNNSITVLEW---TDHTGQENSWKPKKSQAE----KTWTALGAFATDE--YQWL  284

Qy  362  EIDLGEKKKITGIRTTGSTQ-SNFNFYVKS----FVMNFKNNNSKWKIYKGIVNNEEKVFQ  417
Db  285  QIDLNEKKITGIIITTGSTMVSTITMCLPTESCTVMMGRNG----LCTESLVVEQDKIFQ  340

Qy  418  GNSN-FRDPVQNNFIPPIVAR  437
Db  341  GNRKIITRYHRVNNFLPPIIAR  361
```

RESULT 7

```
Q9H2E2
ID   Q9H2E2;      PRELIMINARY;   PRT;   555 AA.
AC   Q9H2E2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
```

Left column

```
RESULT 8
Q9H2E4
ID   Q9H2E4          PRELIMINARY;    PRT;   901 AA.
AC   Q9H2E4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROPILIN-2 SOLUBLE ISOFORM 9.
GN   NRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROSTATIC ADENOCARCINOMA;
RX   MEDLINE=20564205; PubMed=11112349;
RA   Rossignol M., Gagnon M.L., Klagsbrun M.;
RT   "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT   Identification and Distribution of Splice Variants and Soluble
RT   Isoforms.";
RL   Genomics 70:211-222(2000).
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR   EMBL; AF280546; AAG41405.1; -.
DR   HSSP; P12259; 1CZT.
DR   InterPro; IPR000859; CUB.
DR   Pfam; PF00431; CUB; 2.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; -.
SQ   SEQUENCE  555 AA;  62430 MW;  B67B7C992AB0546F CRC64;

Query Match             11.4%;  Score 349;  DB 4;  Length 555;
Best Local Similarity   25.3%;  Pred. No. 2.3e-19;
Matches   109;  Conservative   72;  Mismatches 166;  Indels  84;  Gaps  16;

Qy      88 CGHLVVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
           || :| ||| | |: ||  || | || ||: |||:| |::|| :|| ||||:|| ||:|
Db      28 CGGRLNSKDAGYITSPGTYPQDYPSHQNCEWIVYAPEPNQKIVLNFNHFEIEKHDCKYDF 87

Qy     146 LLF-----TSSSDQYPGYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY-----AS 196
           : |     :||:|  ||:|| :  | ||  || || ||:|:|   | :|| || ::   :|
Db      88 IEIRDGDSESADLLGKHCGNI-APPTISSGSMLYIKFTSDYAROGAGFSLRYEIFKTGS 146

Qy     197 SD-------HPDLITC---LERASHYLKTEYSKF---------------------- 220
           ||       :|:|||    ||:: :| || :|
Db     147 EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPKMEIILLQFLIFDLEHDPLQVGEG 206

Qy     221 -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAITHAGIIADELGCQISVLQRGISRYEG 277
            |   ||||| |:   |||||| ||||   |||:| ||||||||| || ||:|||||||
Db     207 DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSELRSSTGISLITFHTDMAVAK------DG 259

Qy     278 ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE------PDGQIRASSNQSVNESGDQVHWS 333
           : | |:| :: || |:|| ||:|| |: :        |||||||| ||||||| || :|
Db     260 FSARYILVHQEPLENFQ----------CNVPLGMESGRIANEQISASSTY------SDGRWT 305

Qy     334 PEQARLQDQGPSWASGSDSSNHKPREWLEIDLGEKKKITGIRTG--STQSNFNFYVKSF 391
           | |:||  : :|  : :|::  ||  : | | :  |::|| :| |  ||:| |:|||:|
Db     306 PQQSRLHGDDNGWTPNLDSN---KEYLQVDLRFLTMLFAIATQGAISRETQNGTYVKSY 361

Qy     392 VNMFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQONNFIPPIVARYVRVRVPQTWHGSGIA 451
           :|:||||:|||| |||:|||:: ||| |:|:  ::|| ||  |   |  |:||||||:|||| |
Db     362 KLEVSTNGEDMWYRHGKNH--KVFQANNDATEVVLNKLHAPLLTRFVRIRPQTWHSGGIA 419

Qy     452 LKVELIGCQIT 462
           | :|| ||::|
Db     420 LRLELFGCRVT 430
```

Right column

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROPILIN-2B(O).
GN   NRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20564205; PubMed=11112349;
RA   Rossignol M., Gagnon M.L., Klagsbrun M.;
RT   "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT   Identification and Distribution of Splice Variants and Soluble
RT   Isoforms.";
RL   Genomics 70:211-222(2000).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AF280544; AAG41403.1; -.
DR   HSSP; P12259; 1CZT.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR000998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00629; MAM; 1.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; -.
DR   PROSITE; PS00060; MAM_2; 1.
KW   Glycoprotein; Neurone; Receptor; Transmembrane.
SQ   SEQUENCE  901 AA;  101380 MW;  1F3I07A9CC65E5A2 CRC64;

Query Match             11.4%;  Score 349;  DB 4;  Length 901;
Best Local Similarity   25.3%;  Pred. No. 4.8e-19;
Matches   109;  Conservative   72;  Mismatches 166;  Indels  84;  Gaps  16;

Qy      88 CGHLVVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
           || :| ||| | |: ||  || | || ||: |||:| |::|| :|| ||||:|| ||:|
Db      28 CGGRLNSKDAGYITSPGTYPQDYPSHQNCEWIVYAPEPNQKIVLNFNHFEIEKHDCKYDF 87

Qy     146 LLF-----TSSSDQYPGYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY-----AS 196
           : |     :||:|  ||:|| :  | ||  || || ||:|:|   | :|| || ::   :|
Db      88 IEIRDGDSESADLLGKHCGNI-APPTISSGSMLYIKFTSDYAROGAGFSLRYEIFKTGS 146

Qy     197 SD-------HPDLITC---LERASHYLKTEYSKF---------------------- 220
           ||       :|:|||    ||:: :| || :|
Db     147 EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPKMEIILLQFLIFDLEHDPLQVGEG 206

Qy     221 -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAITHAGIIADELGCQISVLQRGISRYEG 277
            |   ||||| |:   |||||| ||||   |||:| ||||||||| || ||:|||||||
Db     207 DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSELRSSTGISLITFHTDMAVAK------DG 259

Qy     278 ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE------PDGQIRASSNQSVNESGDQVHWS 333
           : | |:| :: || |:|| ||:|| |: :        |||||||| ||||||| || :|
Db     260 FSARYILVHQEPLENFQ----------CNVPLGMESGRIANEQISASSTY------SDGRWT 305

Qy     334 PEQARLQDQGPSWASGSDSSNHKPREWLEIDLGEKKKITGIRTG--STQSNFNFYVKSF 391
           | |:||  : :|  : :|::  ||  : | | :  |::|| :| |  ||:| |:|||:|
Db     306 PQQSRLHGDDNGWTPNLDSN---KEYLQVDLRFLTMLFAIATQGAISRETQNGTYVKSY 361

Qy     392 VNMFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQONNFIPPIVARYVRVRVPQTWHGSGIA 451
           :|:||||:|||| |||:|||:: ||| |:|:  ::|| ||  |   |  |:||||||:|||| |
Db     362 KLEVSTNGEDMWYRHGKNH--KVFQANNDATEVVLNKLHAPLLTRFVRIRPQTWHSGGIA 419

Qy     452 LKVELIGCQIT 462
           | :|| ||::|
Db     420 LRLELFGCRVT 430
```

RESULT 9

```
Q9H2D5
ID   Q9H2D5           PRELIMINARY;       PRT;     901 AA.
AC   Q9H2D5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROPILIN-2B(0).
GN   NRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20564205; PubMed=11112349;
RA   Rossignol M., Gagnon M.L., Klagsbrun M.;
RT   "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT   Identification and Distribution of Splice Variants and Soluble
RT   Isoforms.";
RL   Genomics 70:211-222(2000).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AF281074; AAG41899.1; -.
DR   HSSP; P12259; 1CZT.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR004421; FA58_C.
DR   InterPro; IPR009998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00629; MAM; 1.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; UNKNOWN_1.
DR   PROSITE; PS50060; MAM_2; 1.
KW   Glycoprotein; Neurone; Receptor; Transmembrane.
SQ   SEQUENCE   901 AA;  101381 MW;  A131Q304AFD086AB CRC64;
```

```
Query Match             11.4%;  Score 349;  DB 4;  Length 901;
Best Local Similarity   25.3%;  Pred. No. 4.8e-19;
Matches  109;  Conservative  72;  Mismatches 166;  Indels  84;  Gaps  16;

Qy    88  CGHILVTYQDSCTMYMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
             ||   |     :|  ||  : | || |   |  |  ||||  ||| ||||||| ||::
Db    28  CGGRLNSKDAGYITSPGYPQDYPSHQNCEMIVYAPEPNQKIVLNFNPHFEIEKHDCKYDF  87

Qy   146  LLF------TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLIY----AS 196
             |:|      |:|| | |  | | ||||||||:||:||  ||| ::|  | ||:::
Db    88  IEIRDGDSEADLLGKHGNI-APPTISSGSMLYIKFTSDYARQGAGFSLRYEIFKTGS 146

Qy   197  SD------------------------------HPDLITC----LERASHYLKTEYSKF- 220
                                           ||  | |    |:  | || ||||
Db   147  EDCSKNFTSPNGTIESPGFPEKYPNHLDCTFFILAKPKMEILLQFLFIFDLEHDPLQVGEG 206

Qy   221  --CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGCGISVLQRKGISRYEG 277
               ||   |  : |     || |  |:| |  || ||  |:|  ||  || |  | |||   |
Db   207  DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSELRSSTGILSLTFHTDMAVAR------DG 259

Qy   278  ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE-------PDGQIRASSSWQSVNESGDGVHMS 333
             |  | |    | | ||||||||||| ||| |        |||| ||||:|||| ||||:::
Db   260  FSARYIVHQEPLENFQ-------CNVPLGMESGRIANEQISASSTY----SDGRWT 305

Qy   334  PCGAARLQDQGFSWASGCDSSNNHKPREWLEIDLGEKKLITCRTTG--STQSNFHYVVKSF 391
             |||      |   |   |     ||| ||  |:| ||   | |     | :|| | |  |
Db   306  PQQSRLHGDDGWTPNLDSN----KEYLQVDLRFLTMLTAIATQGAISRETQNGYYVKSY 361

Qy   392  VWNFKNNNSKWWTYKGIVNNEEKVPQGNSNFRDPQVQNNFFIPIVARYVRVYPQTWHQRIA 451
             :      |  | |  ||::| | ||||   |  || |  | |  || || || |: | :||
Db   362  KLEVSTNGEDWMVYRHGKNH--KVFQANNDATEVVLNKLHAPLLTRPVRIRRPQTWHSGIA 419
```

RESULT 10

```
Q9H2E3
ID   Q9H2E3           PRELIMINARY;       PRT;     906 AA.
AC   Q9H2E3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROPILIN-2B(5).
GN   NRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20564205; PubMed=11112349;
RA   Rossignol M., Gagnon M.L., Klagsbrun M.;
RT   "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT   Identification and Distribution of Splice Variants and Soluble
RT   Isoforms.";
RL   Genomics 70:211-222(2000).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AF280545; AAG41404.1; -.
DR   HSSP; P12259; 1CZT.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR004421; FA58_C.
DR   InterPro; IPR009998; MAM.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00629; MAM; 1.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C_1; UNKNOWN_1.
DR   PROSITE; PS50060; MAM_2; 1.
KW   Glycoprotein; Neurone; Receptor; Transmembrane.
SQ   SEQUENCE   906 AA;  101955 MW;  0FB01F24D8D9B0BD CRC64;
```

```
Query Match             11.4%;  Score 349;  DB 4;  Length 906;
Best Local Similarity   25.3%;  Pred. No. 4.9e-19;
Matches  109;  Conservative  72;  Mismatches 166;  Indels  84;  Gaps  16;

Qy    88  CGHILVTYQDSCTMYMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
             ||   |     :|  ||  : | || |   |  |  ||||  ||| ||||||| ||::
Db    28  CGGRLNSKDAGYITSPGYPQDYPSHQNCEMIVYAPEPNQKIVLNFNPHFEIEKHDCKYDF  87

Qy   146  LLF------TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLIY----AS 196
             |:|      |:|| | |  | | ||||||||:||:||  ||| ::|  | ||:::
Db    88  IEIRDGDSEADLLGKHGNI-APPTISSGSMLYIKFTSDYARQGAGFSLRYEIFKTGS 146

Qy   197  SD------------------------------HPDLITC----LERASHYLKTEYSKF- 220
                                           ||  | |    |:  | || ||||
Db   147  EDCSKNFTSPNGTIESPGFPEKYPNHLDCTFFILAKPKMEILLQFLFIFDLEHDPLQVGEG 206

Qy   221  --CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGCGISVLQRKGISRYEG 277
               ||   |  : |     || |  |:| |  || ||  |:|  ||  || |  | |||   |
Db   207  DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSELRSSTGILSLTFHTDMAVAR------DG 259

Qy   278  ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE-------PDGQIRASSSWQSVNESGDGVHMS 333
             |  | |    | | ||||||||||| ||| |        |||| ||||:|||| ||||:::
Db   260  FSARYIVHQEPLENFQ-------CNVPLGMESGRIANEQISASSTY----SDGRWT 305

Qy   362  KLEVSTNGEDWMVYRHGKNH--KVFQANNDATEVVLNKLHAPLLTRPVRIRRPQTWHSGIA 419

Qy   420  LRLELFGCRVT 430
```

```
Db   452  LKVELCLGCQIT 462
```

```
RESULT 11
ID   Q9H2D4
AC   Q9H2D4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)                           PRELIMINARY;      PRT;   906 AA.
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROPILIN-2B(S).
GN   NRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20564205; PubMed=11112349;
RA   Rossignol M., Gagnon M.L., Klagsbrun M.;
RT   "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT   Identification and Distribution of Splice Variants and Soluble
RT   Isoforms.";
RL   Genomics 70:211-222(2000).
EMBL; AF281074; AAG41900.1; -.
HSSP; P12259; 1CZT.
InterPro; IPR000859; CUB.
InterPro; IPR000421; FA58.
-!-  SIMILARITY: CONTAINS 2 CUB DOMAINS.
-!-  SIMILARITY: CONTAINS 1 MAM DOMAIN.
-!-  SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
-!-  SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
InterPro; IPR000998; MAM.
Pfam; PF00431; CUB; 2.
Pfam; PF00629; MAM; 1.
PRINTS; PR00020; MAMDOMAIN.
SMART; SM00042; CUB; 2.
SMART; SM00231; FA58C; 2.
SMART; SM00137; MAM; 1.
PROSITE; PS01180; CUB; 2.
PROSITE; PS01285; FA58C_1; 2.
PROSITE; PS50060; MAM_2; 1.
PROSITE; PS00060; UNKNOWN_1.
KW   Glycoprotein; Neurone; Receptor; Transmembrane.
SQ   SEQUENCE  906 AA;  101956 MW;  EB183F265A57B089 CRC64;

Qy     260 FSARYLVHQEPLENFQ---------CNVPLGMESGRIANEQISASSTY-------SDGRWT   305
               |: |||| ||  ||      :|  |: | |||   |::|| ||     ||| |||:||
Db     334 PGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTTG--STQSNFNFYVKSF   391

Qy     306 PQQSRLHGDDNGWTPNLDSN----KEYLQVDLRFLTMLTAIATQGAISRETQNGIYVKSY   361
               |||  ||| |||  ||     :||||||:||:|| || | |: ||||:|| ||:|||
Db     392 VNNEFKNNNSKWTYKGIVNNEEKVEFQGNSNFRDPVQQNFPIPIVARYVRVRPQTWHGIA   451

Qy     362 KLEVSTNGEDWMVYRHGKNH--KVFQANNDAIEVVLNKLHAPLLTRFVRIRPQTWHSGIA   419
               |||||||||||||||||   ||||||||||||||:||||||||||||||||||||||
Db     452 LKVELLGCQITT                                                 462

Qy     420 LRIELFGCRVT   430
               |||||||||||
Db     420 LRIELFGCRVT   430
```

Query Match　　11.4%;　　Score 349;　DB 4;　　Length 906;
Best Local Similarity 25.3%;　Pred No. 4.9e-19;
Matches 109;　Conservative 72;　Mismatches 166;　Indels 84;　Gaps 16;

```
Qy      88 CGHLVTYQDSGTMTSKNYPGTIYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY   145
               ::  |:   ||  |  |||  | ::| |  |:      |  : | | |:|| ||| :|
Db      88 IEIRDGDSESADLLGKHGCNI-APPTLISSGSMYLIKFTSDYARQGAGFSLRYEIFKKTGS   146

Qy     146 LIF-----TSSSDQYGPYCCGSMTVFKELLLNTSEVTVRFESGSHISGRGFLLTY-----AS   196
                |       ||: | |||::|: |  | || |   |:||| |:|||||||| ||
Db     147 EDCSKNFISPNGTIESPGFPEKXYPHNLDCTFTILAKPKMEIILLQFLLFDLEHDPLQVEG   206

Qy     197 SD-------------------------HPDLITC---LERASHYLKTEYSKF-----------   220
               ::                           :|||    | ||: :||:|:|
Db
```

```
RESULT 12
ID   Q96I90
AC   Q96I90;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)                           PRELIMINARY;      PRT;   609 AA.
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO NEUROPILIN 1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;
RA   Strausberg R.;
RT   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RL   EMBL; BC007737; AAH07737.1; -.
SQ   SEQUENCE  609 AA;  68347 MW;  798AAFD2D568C127 CRC64;

Qy     221 -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIIHAGLIADELGGGISVLQRVSRYEG   277
               |   |: :| |:    || | |: ||::| || : : || : |   |:: |:::| |
Db     207 DCKYDWLDITWDGIPHVGPLIGKYCCGTKIPSELRSSTGITCLSLSETFHTDMAVAK--   259

Qy     278 LLANGVLSRDGSLSDKRFLFTSNGCGSRSLSFE-----PDGQIRASSSWQSVNESGCDVVHMS   333
               ||  |: |||| ||||||| ||||| | ||| :       || | ||:   || |  :|:|
Db     260 FSARYLVHQEPLENFQ------CNVPLGMESGRIANEQISASSTY------SDGRWT   305

Qy     334 PGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTTG--STQSNFNFYVKSF   391
               ||||||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db     334 PGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTTG--STQSNFNFYVKSF   391

Qy     392 VNNEFKNNNSKWTYKGIVNNEEKVEFQGNSNFRDPVQQNFPIPIVARYVRVRPQTWHGIA   451
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 VNNEFKNNNSKWTYKGIVNNEEKVEFQGNSNFRDPVQQNFPIPIVARYVRVRPQTWHGIA   451

Qy     452 LKVELLGCQITT                                                 462
               ||||||||||||
Db     452 LKVELLGCQITT                                                 462

Qy     420 LRIELFGCRVT   430
               |||||||||||
Db     420 LRIELFGCRVT   430
```

Query Match　　11.0%;　　Score 335;　DB 4;　　Length 609;
Best Local Similarity 24.9%;　Pred No. 3.5e-18;
Matches 124;　Conservative 76;　Mismatches 176;　Indels 122;　Gaps 23;

```
Qy      64 RGL-----LALILAVSAPLRLQAAFELGDGCGHLVTYQDSGTMTSKNYPGTIYPNHTVCEK   117
               ||      |  | | :|   | |    | ||| |||  |||| |  |||| |||| ||||
Db       3 RGLPLLCAVLALVLAPAGAFR------NDKCGDTIKIESPGYLIFSPGYHSYHSGERCEW   56

Qy     118 TITVPKG-KRLILRLG-DLDIESQTCASDYL-LFTSSSDQ---YGPYCCGSMTVFKELLLN   171
               |||  |  |||||| | |||||| ||||| |  |||| |     | |||| ||| ||||
Db      57 LIQAPDPYQRIMININFNPHFDLEDRCDCKYDYVEVFDGENENGHFRGKFCGKI-APPVVSS   115

Qy     172 TSEVTVRFESGSHISGRGFLLTYA---------------SSDHIPDLITC---   205
               :|  |  |   : |: || | ||                  || |  ||
Db     116 GPFLIFKVFSDYETGHGAGFSIRYEIFKRGPECSQNYTTPSGVIKSPGFPEKYPNSLECTY   175

Qy     206 IVFAPKMSEILLEFESFPLEPDSNP--PGGMFCRYDRLEI-----VGPH   253
               | |||   |      | || ||     ||  |||| |||        :||
Db     176 LERASHYLKTEYSKFPCPAG---CRDVAGDISGNNVDGYRDTSLLCRAAH   252

Qy     253 AG-IIADELGGQI---SVLQRKGISRTEGILANCYLSRDGSLSDKRFLF   297
               :| || |    |    || |:     |  | | ||| ||| |||||||
Db     253 IGRYCGQKTRPGRLRSSSGILSMVFTDSAIAKEGFSANYSVLQSSY-SEDFK   274

Qy     298 TSNGCGSRSLSFEP----DGQIRASSSWQSVNESGCDVVHHSPGQARLQDQGPSWASGDSSN   353
               ::|||| | | :|     |  |||| |  ||||| | |  || ||||||||||||| |:::
Db     275 ----CMEALGMESGRIHSDQITPASSSQYST-----IWSAEKSRLNTYPENGWTPGGDSY   322

Qy     354 NHKPREWLEIDLGEKKKITGIRTTG---STQSNFNFYVKSFVWNEFKNNNSKWTYKGIVNN   411
               ||||||||||||||||||||||||    ||||||||| |||| |||||||||||||||||
Db
```

RESULT 13

```
Q96IH5

ID   Q96IH5      PRELIMINARY;    PRT;   644 AA.
AC   Q96IH5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROPILIN 1 (Human).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY, AND RENAL CELL ADENOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007533; AAH07533.1; -.
SQ   SEQUENCE   644 AA;  71907 MW;  9204B6F5C3CD2A21 CRC64;
```

```
Query Match            24.9%;  Score 335;  DB 4;  Length 644;
Best Local Similarity  24.9%;  Pred. No. 3.8e-18;
Matches  124;  Conservative   76;  Mismatches  176;  Indels  122;  Gaps  23;

Qy     64 RGL------LALLLAVSAPLRLQAEELGDGCGHLVTVQDSGTMTSKNYPGTYPNHTVCEK 117
              |||     |:||::|||  ||   || :   :::|
Db      3 RGLPLCAVLALVLAPAGAFR-----NDKCGDPTIKIESPGYLTSPGYPHSYHPSEKCEW  56

Qy    118 TIIVPKG-KRLIRLG-DLDIESQTCASDYL-LFTSSSDQ----YGPYCGSMTVPKELLLN 171
Db     57 LIQAPDPYQRIMINFNPHFDLEDRDCKYDYVEVFDGENENGHFRGKFCGKI-APPVVSS  115

Qy    172 TSEVTVRFESGSHISGRGFLLTYA----------------SSDHPDLITC--        205
Db    116 GPFLFIKVSDYETHGAGFSIRYEIFKRGPECSQNYTTPSGVIKSPGFPEKYPNSLECTY  175

Qy    206 --------LERASHYLKTEYSKFCPAG---CRDVAGDIGSNMVDGYRDTSLLCKAAIH  252
Db    176 IVFAPKMSEIILEFESFDLEPDSNP--PGGMFCRYDRLEI-----VGPH          223

Qy    253 AG-IIADELGGQI-------SVLQRKGISRYEGIILANGVLSRDGSLSDKRFLE      297
Db    224 IGRYCGQKTPGRIRSSSGILSMVFYTDSAIAKEGFSANYSVLQSSV-SEDFK-----   274

Qy    298 TSNGCSRSLSFEP------DGQIRASSMWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSN 353
Db    275 ----CMEALGMESGEIHSDQITASSQYST---------NWSAERSRLNYPENGWTPGEDSY 322

Qy    354 NHKPREWLEIDLGEKKKIITGIRTG---STQSNFNEYVKSFYVNNFKNNSKWKTYGIVNN 411
Db    323 ----REWIQVDLGLLRFVTAVGTQGAISKETKKKYYVKTYKIDVSSNGEDWITIK--EGN 376

Qy    412 EEKVFQGNSNRFRDPVQNNFIPPIVARYVRVPQTWHQRIALKVELIGCQITT----QGNDS 467
Db    377 KPVLFQGNNTNPFDVVAVFPKPLITRFVRIKPATWETGISMRFEVYGCKITDYPCSGMLG 436

Qy    468 LVWRKTSQSTSVSTKKED 485
Db    437 MVSGLISDSQOTTSSNGQD 454
```

RESULT 14

```
Q9H2E1

ID   Q9H2E1      PRELIMINARY;    PRT;   704 AA.
AC   Q9H2E1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROPILIN-1 SOLUBLE ISOFORM 11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROSTATIC ADENOCARCINOMA;
RX   MEDLINE=20564205; PubMed=11112349;
RA   Rossignol M., Gagnon M.L., Klagsbrun M.;
RT   "Genomic Organization of Human Neuropilin-1 and Neuropilin-2 Genes:
RT   Identification and Distribution of Splice Variants and Soluble
RT   Isoforms.";
RL   Genomics 70:211-222(2000).
CC   -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
DR   EMBL; AF280547; AAG41406.1; -.
DR   HSSP; P12259; 1C2T.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR000421; FA58_C.
DR   Pfam; PF00431; CUB; 2.
DR   Pfam; PF00754; FA58C; 2.
DR   SMART; SM00042; CUB; 2.
DR   SMART; SM00231; FA58C; 2.
DR   PROSITE; PS01180; CUB; 2.
DR   PROSITE; PS01285; FA58C; 1.
SQ   SEQUENCE   704 AA;  79002 MW;  1625DDD561459307 CRC64;
```

```
Query Match            24.9%;  Score 335;  DB 4;  Length 704;
Best Local Similarity  24.9%;  Pred. No. 4.4e-18;
Matches  124;  Conservative   76;  Mismatches  176;  Indels  122;  Gaps  23;

Qy     64 RGL------LALLLAVSAPLRLQAEELGDGCGHLVTVQDSGTMTSKNYPGTYPNHTVCEK 117
Db      3 RGLPLCAVLALVLAPAGAFR-----NDKCGDPTIKIESPGYLTSPGYPHSYHPSEKCEW  56

Qy    118 TIIVPKG-KRLIRLG-DLDIESQTCASDYL-LFTSSSDQ----YGPYCGSMTVPKELLLN 171
Db     57 LIQAPDPYQRIMINFNPHFDLEDRDCKYDYVEVFDGENENGHFRGKFCGKI-APPVVSS  115

Qy    172 TSEVTVRFESGSHISGRGFLLTYA----------------SSDHPDLITC--        205
Db    116 GPFLFIKVSDYETHGAGFSIRYEIFKRGPECSQNYTTPSGVIKSPGFPEKYPNSLECTY  175

Qy    206 --------LERASHYLKTEYSKFCPAG---CRDVAGDIGSNMVDGYRDTSLLCKAAIH  252
Db    176 IVFAPKMSEIILEFESFDLEPDSNP--PGGMFCRYDRLEI-----VGPH          223

Qy    253 AG-IIADELGGQI-------SVLQRKGISRYEGIILANGVLSRDGSLSDKRFLE      297
Db    224 IGRYCGQKTPGRIRSSSGILSMVFYTDSAIAKEGFSANYSVLQSSV-SEDFK-----   274

Qy    298 TSNGCSRSLSFEP------DGQIRASSMWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSN 353
Db    275 ----CMEALGMESGEIHSDQITASSQYST---------NWSAERSRLNYPENGWTPGEDSY 322

Qy    354 NHKPREWLEIDLGEKKKIITGIRTG---STQSNFNEYVKSFYVNNFKNNSKWKTYGIVNN 411
Db    323 ----REWIQVDLGLLRFVTAVGTQGAISKETKKKYYVKTYKIDVSSNGEDWITIK--EGN 376

Qy    412 EEKVFQGNSNRFRDPVQNNFIPPIVARYVRVPQTWHQRIALKVELIGCQITT----QGNDS 467
Db    377 KPVLFQGNNTNPFDVVAVFPKPLITRFVRIKPATWETGISMRFEVYGCKITDYPCSGMLG 436

Qy    468 LVWRKTSQSTSVSTKKED 485
Db    437 MVSGLISDSQOTTSSNGQD 454
```

```
Db   437 MVSGLISDSQITSSNQGD 454

RESULT 15

ID      Q90X38        PRELIMINARY;        PRT;      921 AA.
AC      Q90X38;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      NEUROPILIN.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97433084; PubMed=9287753;
RA      He Z., Tessier-Lavigne M.;
RT      "Neuropilin is a receptor for the axonal chemorepellent Semaphorin
RT      III.";
RL      Cell 90:739-751(1997).
CC      -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC      -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC      -!- SIMILARITY: BELONGS TO THE NEUROPILIN FAMILY.
CC      -!- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC      -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR      EMBL; AF018957; AAC53345.1; -.
DR      HSSP; P12259; 1C2T.
DR      InterPro; IPR000859; CUB.
DR      InterPro; IPR000421; FA58_C.
DR      InterPro; IPR011092; HLH_dim.
DR      InterPro; IPR000998; MAM.
DR      Pfam; PF00431; CUB; 2.
DR      Pfam; PF00754; F5_F8_type_C; 2.
DR      Pfam; PF00629; MAM; 1.
DR      PRINTS; PR00020; MAMDOMAIN.
DR      SMART; SM00042; CUB; 2.
DR      SMART; SM00231; FA58C; 2.
DR      SMART; SM00137; MAM; 1.
DR      PROSITE; PS01180; CUB; 2.
DR      PROSITE; PS01285; FA58C_1; 2.
DR      PROSITE; PS01286; FA58C_2; 2.
DR      PROSITE; PS00740; MAM_1; 1.
DR      PROSITE; PS00741; MAM_2; 1.
DR      PROSITE; PS50060; MAM_2; 1.
KW      Glycoprotein; Neurone; Receptor; Transmembrane.
SQ      SEQUENCE 921 AA; 103052 MW; 58B29A9AA4978971 CRC64;

Query Match        10.7%;   Score 325.5;   DB 11;   Length 921;
Best Local Similarity   24.0%;   Pred. No. 3.8e-17;
Matches  110;   Conservative  81;   Mismatches 173;   Indels  95;   Gaps  18;

Qy    64 RGLLALILAVSAPLRLQAEELGDGCGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK 123
          ||||  | |:  | | |:  |  ::| ||:::| | : |: | ||||||:::||:|:||
Db     3 RGLPLCATLALALALGAFR-SDKCGGTKIENPGYLTSPGYPHSYHPSERCEWLIQAPE 61

Qy   124 G-KRLILRLG-DLDIESQTCASDYLFTSSSDG----YGPYCGSMTVPKELLLNTSEVTV 177
          | :|||| |  :|| ::|||| || |:||:       |||| |:|  || ||:||: |
Db    62 PYQRIMINFNPHPFDLEDRDCKYDIVEVIDGENEDGRLWGKPFCGKI-APSPVVSSGPFLPI 120

Qy   178 RFESGSHISGRGFLLTYA--------SSDHPPLLTTC------ 205
          :|:|: : | :|| ||             |:|| | ||
Db   121 KPVSDYETHGAGFSIRYEIFKRGPECSQNYTAPTGVIKSPGFPPEKYPNSLECTYIIFAPK 180

Qy   206 -----LERASHYLKTEYSKPCPAG--CRDVAGDISGNMVDGYRDTSLLCKAAIHAGIAD 258
              || |:||: :::| : |    ::  | ||:| : | : :| | : | | | | |
Db   181 MSEIILPFESFDIEQDSNP--PGGMFCRYDRLEI----WIGPPE------------VGP 221

Qy   259 ELGGQISVLQRKGISRYEGILA----------NGVLSRDGSLSDKRFLFTSNG----CSRSLSFE 309
           | | |:| |::| ::||                | :| | |         |::
Db   222 HIGRYCGQKTPGRIRSSSGILSMVFYDSAIAKEGFSANYSVLQSSISEDFKCMEALGME 281

Qy   310 P----DGGIRASSSWQSVWNESGDQVHWSPGGQARLQDGGPSWASGDSSWNHKPREWLEIDL 365
          |      ||::|   : |  |: | | | | | : :  | || | | | | ::|  |:||||
Db   282 SGEIHSDQIITASSQYGT-------NWSVERSRLNTPENGWTPGEDSY---REWIQVDL 329

Qy   366 GEKKIITGIIRTTG---STQSNFNTYKSFPVMNFKNNNSKKYTTKGIYNNEEKVFFQGNSNFR 423
          |:  ||  | :| |   :|:|  :| |: | | |  || ||  || || || |  | :|:| :
Db   330 GLLRPVTAVGTQGAISKETKKKYVKTVRVDISSNGEDWITLK--EGNKAIIFQGNTNPT 387

Qy   424 DPVQNNEIIPPIVARVVRVPQTWHQRIALKVELLTGCQIT 462
          || :|||| | :|  |:|  | | ||||| ||| | | :|
Db   388 DVVFGVFPKPLITRFVRIKPASWETGISMRFEVYGCKII 426
```

Search completed: July 9, 2002, 02:25:21
Job time: 678 sec

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: July 8, 2002, 23:32:43 ; Search time 3285.5 Seconds
(without alignments)
11216.445 Million cell updates/sec

Title: US-09-691-344A-3
Perfect score: 1761
Sequence: 1 atgggattcggtcgcgggca.........caagtgatatggcaggttaa 1761
Scoring table: IDENTITY_NUC
   Gapop 10.0 , Gapext 1.0
Searched: 1797656 seqs, 1046326293 residues
Total number of hits satisfying chosen parameters: 3595312
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0% Maximum Match 100%
   Listing first 45 summaries

Database :   GenEmbl:*

No.	Database
1:	gb_ba:*
2:	gb_htg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_cv:*
6:	gb_pat:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vi:*
15:	gb_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pat:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_sts:*
28:	em_un:*
29:	em_vi:*
30:	em_htg_hum:*
31:	em_htg_inv:*
32:	em_htg_other:*
33:	em_htg_inv:*

SUMMARIES

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1761	100.0	1761	1761	6	AX118818 Sequence
2	1620	92.0	1620	1620	6	AX118820 Sequence
3	1522	86.4	1522	1768	6	AX118822 Sequence
4	1464	83.1	1464	1464	6	AX118816 Sequence
5	359.6	20.4	257	249429	2	AC099419 Rattus no
6	257	14.6	257	131060	2	AC099419 Rattus no
7	229.8	13.0	2642		10	AF387548 Mus muscu
8	225.8	12.8	2020		9	AK055350 Homo sapi
9	224.2	12.7	2939		6	AF387547 Homo sapi
10	220.4	12.5	2339		6	AF387549 Homo sapi
11	218	12.4	142616		2	HS94G16 Human DNA s
12	218	12.4	142616		2	AC026646 Homo sapi
13	159	9.0	117300		2	AC099428 Rattus no
14	134.4	7.6	142616		2	AC026646 Homo sapi
15	132.8	7.5	1388		6	HUMCUB1 Human mRNA
16	131.4	7.5	2072		9	HSA420528 / AJ420528 Homo sapl
17	95	5.4	6895		4	BOVFACV2A Bos taurus
18	95	5.4	6910		4	BOVFACV Bos taurus
19	94.4	5.4	6893		4	HUMFVA Human coagu
20	92.8	5.3	6909		6	M81440 Bos taurus
21	92.8	5.3	6909		6	M81441 Bos taurus
22	92.8	5.3	6909		6	M14335 Human coagu
23	92.8	5.3	6909		6	A46255 Sequence 1
24	92.8	5.3	6909		6	A63218 Sequence 1
25	92.8	5.3	6909		6	AR071379 Sequence
26	88	5.0	7062		4	AX146885 Sequence
27	83.2	4.7	7145		4	E29754 Method for
28	81.6	4.6	7032		4	HUMFAV Human coagu
29	81.6	4.6	7032		4	AF191308 Sus scrof
30	80.6	4.6	6585		4	M16967 Human coagu
31	80.6	4.6	8720		6	AF191308 Sus scrof
32	80	4.5	10698		6	AF049489 Canis fam
33	80	4.5	2493		9	AF049489 Canis fam
34	78.4	4.5	2536		9	AF016234 Canis fam
35	78.4	4.5	4278		9	AR159794 Sequence
36	78.4	4.5	4281		9	U52225 Mus musculu
37	78.4	4.5	4551		9	MMUS2925 Mus musculu
38	78.4	4.5	4670		9	AX253510 Sequence
39	78.4	4.5	4551		9	AX253521 Sequence
40	78.4	4.5	4670		9	M90707 Homo sapien
41	78.4	4.5	4999		9	HUMF8C Homo sapien
42	78.4	4.5	4999		9	BC022513 Homo sapien
43	78.4	4.5	5035		9	I0644 Sequence 4
44	78.4	4.5	5035		9	I0643 Sequence 3
45	78.4	4.5	6996		6	I0641 Sequence 1

ALIGNMENTS

RESULT 1
AX118818

LOCUS   AX118818   1761 bp   DNA   linear   PAT 11-MAY-2001
DEFINITION   Sequence 3 from Patent WO0129219.
ACCESSION   AX118818
VERSION   AX118818.1   GI:14035775
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1761)
AUTHORS   Donoho,G., Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and
   Sands,A.T.
TITLE   Human cub-domain-containing protein and gene encoding the same
JOURNAL   Patent: WO 0129219-A 3 26-APR-2001;
   Lexicon Genetics Incorporated (US)
FEATURES   Location/Qualifiers
   source   1..1761
     /organism="Homo sapiens"
     /db_xref="taxon:9606"
BASE COUNT   471 a   402 c   483 g   405 t
ORIGIN

```
Query Match                100.0%;  Score 1761;  DB 6;  Length 1761;
Best Local Similarity      100.0%;  Pred. No.: 0;  Matches 1761;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
Qy      1 atggggatccgttcgggggcaaggcggcagcggcgccggctccgcaggg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGGGATCCGTTCGGGGGCAAGGCGGCAGCGGCGCCGGCTCCGCAGAG 60

Qy     61 gaggcggggcgcccagccggccggccgggcagcggggcccgccggccggcggccgctgcc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GAGGCGGGGCGCCCAGCCGGCCGGCCGGGCAGCGGGGCCCGCCGGCCGGCGGCCGCTGCC 120

Qy    121 aagcggggcgcccggggggtcaatggcgcagctgcggctcggatccgtcgaggggcacttggccggg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGCGGGGCGCCCGGGGGGTCAGCTGCGGCTCGGATCCGTCGAGGGGCAGCTTGCC 180

Qy    181 gctgcggcggggcggggggctctggcctggctgtctgtgctttgtctgtctgcccgggggctgcggg 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCTGCGGCGGGGCGGGGGGCTCTGGCCTGGCTGTCTGTGCTTTGTCTGTCTGCCCGGGGCTGCGGG 240

Qy    241 gcgggaggggagctgggtgtgggtgaccctggtcacaggtccctgggtggcacaatg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCGGGAGGGGAGCTGGGTGTGGGTGACCCTGGTCACAGGTCCCTGGGTGGCACAATG 300

Qy    301 acaatctaagaatatccgggggacctaccccaatacaacactgttactggacaaattaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACAATCTAAGAATATACCGGGGACCTACCCCAATACAACACTGTTACTGGACAAATTACA 360

Qy    361 gtaccaaagggggaaaaagactcattctgagggttgggggatatcgatcccccagacc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTACCAAAGGGGGAAAAAGACTCATTCTGAGGGTTGGGGGATATCGAATCCCCAGACC 420

Qy    421 tgtgctctcgactatctcctctctcccacctctcgaggttgggcacaatg 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGTGCTCTTCGACTATCTCTCTCTCTCACCTCTCGAGGTTGGGCACAATG 480

Qy    481 agtatgaagactgttccccaaaagaacatctgttgatacacaagctgaaagtaaccctccgtcttgag 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGTATGAAGACTGTTCCCCAAAGAACATCTGTTGATACACAAGTAACCCTTCCGTCTTGAG 540

Qy    541 agtggatcccaacattctgccgggggtttttgctggacctatgccagcgagcacatcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGTGGATCCCACATTCTGCCGGGGGTTTTTGCTGGACCTATGCCAGCGAGCACATCCA 600

Qy    601 gattaataacatgtctggaacgagcagctagagacagagaatccagcaaatcc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GATTAATAACATGTTTGGAACGAGCAGCTAGAGACAGAGAATCCAGCAAATCC 660

Qy    661 tgcccaagctggttctgtgtagagacgtagcaagaccatttctgggaataatatggctgatga 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TGCCCAAGCTGGTTCTGTAGAGACGTAGCACGAGGAGCATTTCTGGGAATATGGTGATGATGA 720

Qy    721 agagatacctccttattgtgcaaagctgccatccatcatcgcaggaataattgctgtgaacta 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGAGATACCTCTTTATTGTGCAAAGCTGCCATCCATCATCGCAGGAATAATTGCTGTGAACTA 780

Qy    781 ggtggccagatcagtgtgctcagcgagcaatcgatcgtatgtaaggattctgtcgcc 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGTGGCCAGATCAGTGTGCTCAGCGGGCAAAATCAGTCAGATTGTAAGGGATTCTGTGCC 840

Qy    841 aatggtcttctcgaggatggtggttccctgtcagcaagcaagcaggatcctgttacctccaat 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AATGGTGTTCTTCGAGGATGGTGGTTCCCTGTCAGCAAGCAAGCAGGATTCTGTTTACTCCAAT 900

Qy    901 ggtggcagagatccgacggcaaaatcagcaagctcttcctcatgg 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GGTGGCAGCAGAGATCCGACGGCGAAAATCAGCAAGCTCTTCCTCATGG 960

Qy    961 caatccgtcaatgaagtgaggtcagaagttcactgctcctcctggccaagcccgacttcag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CAGTCCGTCAATGAAGTGAGGTTCACTGCTCTCCTGGCCAAGCCCGACTTCAG 1020
```

```
Qy   1021 gaccaagggccatcatcgggcttcgggcgcgacgtgctagcgaacacaaccacacagcagagtgg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GACCAAGAGCCCATCCATCGGGCTTCCGGCGGCAGTGTAGCAACAACCACACAGCAGAGTGG 1080

Qy   1081 ctggagatcgatcttgggggaggaagaaagaaaaaacaaccgggacgatctaca 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTGGAGATCGATCTTGGGGGGAGGAAGAAAGAAAAAACAACCGGGACGATCTACA 1140

Qy   1141 cagtccgaactcaactctatatgtgaataatgaaaaagttttgtcgaacttcaaaacatcaattct 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CAGTCGGAACTTCAACTTTATATGTGATAATGAAAAAGTTTGTGATGAACTTCAAAACAATAATTCT 1200

Qy   1201 aagtggaagaacctataaaagaattcgtgaataatgaagaaaaagttgttcatgggtaactct 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AAGTGGAAGAACCTTCAACTTTATGTGAATAATGAAGAAAAAGGTGTTCATGGGTAACTCT 1260

Qy   1261 aacttctgggaaccagtgcaaaacaacaatttcatcccctccctgcagatatgtcgcgg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AACTTTCTGGGACCCAGTTGCAAAAACAATTTCATCCCCTCCCTGCAGATATGTGCGGG 1320

Qy   1321 gttgtccccagacatggcaaccagggaggagttagctctgagggtgtcatcttggttcccag 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GTTGTCCCCAGACATGGCACCAGGAGGAGATAACCTTGAAGGTGGAGCCTCATTGGTTGCCAG 1380

Qy   1381 attacacaaggtaatgattcattggtgtggcgcaagaacaagtcatccaagtgttca 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ATTACACAAGGTAATGATTCATTGGTGTGGCGCAAGACACAAGTCATCAGCCAGTGTTCA 1440

Qy   1441 actaagaaagaagatgagacatcacaagccaatcccctccccgaagaacatcccacagga 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ACTAAGAAAGAAGATGAGACATCACAAGCCAATCCCCTCCCCGGAGAAACATCCCACAGGA 1500

Qy   1501 ataacattacaacggttggctattccaatggtcactgtgga 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ATAAACATTACAACGGTTGGCTATTCCAATGGTCACTGTGGA 1560

Qy   1561 atggggatctttgcagccctttagaagagaagaagaagaagcaagtccgtatggatcaagg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATGGGGATCTTTGCAGCCCTTTACAGAAGAGAAGAAGAAGTCCGTATGGATCAGCC 1620

Qy   1621 gaggctcagaaaacagagactgttggaagacagagatataatatccccttttgccgagacatcagtca 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GAGGCTCAGAAAACAGACTGTTGGAAGACAGAGATTAAATATCCCTTTGCCGACATCAGTCA 1680

Qy   1681 gctgagttaccatcagctatgatcatgagatgaggagagatgacacaaagagttagatctcatc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GCTGAGTTACCATCAGCTATGATCATGAGATGAGGAGAGATGACACAAAGAGTTAGAGTCTCATC 1740

Qy   1741 acaagtcgatatggcagggttaa 1761
          ||||||||||||||||||||||||||||
Db   1741 ACAAGTCGATATGGCAGGGTTAA 1761
```

```
RESULT   2
LOCUS       AX118820                    1620 bp    DNA        linear        PAT 11-MAY-2001
DEFINITION  AX118820 Sequence 5 from Patent WO0129219.
ACCESSION   AX118820
VERSION     AX118820.1  GI:14035776
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1620)
  AUTHORS   Donoho,G., Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and
            Sands,A.T.
  TITLE     Human cub-domain-containing protein and gene encoding the same
  JOURNAL   Patent: WO 0129219-A 5 26-APR-2001;
            Lexicon Genetics Incorporated (US)
FEATURES             Location/Qualifiers
     source          1..1620
```

BASE COUNT   456 a   353 c   422 g   389 t
ORIGIN

/organism="Homo sapiens"
/db_xref="taxon:9606"

Query Match   92.0%;   Score 1620;   DB 6;   No. 2 2e-299;   Length 1620;
Best Local Similarity   100.0%; Pred.
Matches 1620;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   142 atggtgccgcgcacccccggcgcggcggcggcacactggtcgcggggctgtcgcggggcctccctg 201
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 atggtgccgcgcacccccggcgcggcggcggcacactggtcgcggggctgtcgcggggcctccctg   60

Qy   202 gccttgctgcctcgccgccccggccgcctcggccgcctccggctccggctcggggtgtatggc 261
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 gccttgctgcctcgccgccccggccgcctcggccgcctccggctccggctcggggtgtatggc 120

Qy   262 tgtggacacctgatgactattcaaggatagtggcacatgacatctaagaacatatcccggg 321
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 tgtggacacctgatgactattcaaggatagtggcacatgacatctaagaacatatcccggg 180

Qy   322 acctaccccaatcacactgtttgcggaaaaagaaaattacaccggg 381
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   181 acctaccccaatcacactgtttgcggaaaaagaaaattacaccggg 240

Qy   382 atcctgaggtgtggagagttgggatatcgaatccatagcatactgtgttgcttcgactatctctc 441
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 atcctgaggtgtggagagttgggatatcgaatccatagcatactgtgttgcttcgactatctctc 300

Qy   442 ttcacagctcttcagatcaaatatggttcccatacttgtatcgaatccccaaagaa 501
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ttcacagctcttcagatcaaatatggttcccatacttgtatcgaatccccaaagaa 360

Qy   502 ctcttgttgaacaacaagtccgtctggaagtggatgcacagtggagtatccccacattctgcc 561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ctcttgttgaacaacaagtccgtctggaagtggatgcacagtggagtatccccacattctgcc 420

Qy   562 cggggttcttgctgatcctatgcagcgagcagccaccaccaagatttaatacatgttttggaa 621
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 cggggttcttgctgatcctatgcagcgagcagccaccaccaagatttaatacatgttttggaa 480

Qy   622 cggactagccattcttgaagcacaagacaaatctgcccagctgggttgataagagac 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 cggactagccattcttgaagcacaagacaaatctgcccagctgggttgataagagac 540

Qy   682 gtagcagggagacaatttcttggagaatgatagagagataacacttctttattctgtgtgc 741
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 gtagcagggagacaatttcttggagaatgatagagagataacacttctttattctgtgtgc 600

Qy   742 aaagctgccatccatgcaggagaatattgctgatgaactagtagtggccagatagtgtgctt 801
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 aaagctgccatccatgcaggagaatattgctgatgaactagtagtggccagatagtgtgctt 660

Qy   802 cagcgacaaagggatcagtcgatgatgaggagggattcggccaatggtgtcttttcggaggat 861
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 cagcgacaaagggatcagtcgatgatgaggagggattcggccaatggtgtcttttcggaggat 720

Qy   862 ggtccctgtcagacaagcgaattctgttacctccaatggttgcaagaagatcctgagt 921
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ggttccctgtcagacaagcaagcgaattctcgtttacctccaatggttgcagagatcctgagt 780

Qy   922 ttgaacctgcaggccaaatcagagctctctcctactggcagtcgtcaagtggccaataga 981
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 tttgaacctgcaggccaaatcagagctctctcctactggcagtcgtcaagtggacaata 840

Qy   982 gaccaagttcactggtctctggccaagcccaccatcatgggct 1041
          |||||||||||||||||||||||||||||||||||||||||||
Db   841 gaccaagttcactggtctctggccaagcccaccatcatgggct 900

Qy 1042 tcgggcgacagtagcacaacacaacacggagggtggctcggagatcgattggggggag 1101
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 tcgggcgacagtagcacaacacaacacggagggtggctcggagatcgattggggggag 960

RESULT 3

LOCUS   AX118822   1768 bp   DNA   linear   PAT 11-MAY-2001
DEFINITION   Sequence 7 from Patent WO0129219.
ACCESSION   AX118822
VERSION   AX118822.1 GI:14035777
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1768)
  AUTHORS   Donoho,G., Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and
             Sands,A.T.
  TITLE   Human cub-domain-containing protein and gene encoding the same.
  JOURNAL   Patent: WO 0129219-A 7 26-APR-2001;
             Lexicon Genetics Incorporated (US)
FEATURES   Location/Qualifiers
     source   1..1768
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   539 a   370 c   421 g   438 t
ORIGIN

Query Match   86.4%;   Score 1522;   DB 6;   No. 1.2e-280;   Length 1768;
Best Local Similarity   100.0%; Pred.

Qy 1102 aaaagaaaataacaggaattaggaccacaagtactctacacagtggaagaacttcaactttttat 1161
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 aaaagaaaataacaggaattaggaccacaagtactctacacagtggaagaacttcaactttttat 1020

Qy 1162 gttaagagttttgtgatgaactccaaaacaataattctaacacagctgaagaccatataaagga 1221
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 gttaagagttttgtgatgaactccaaaacaataattctaacacagctgaagaccatataaagga 1080

Qy 1222 attgtgaataatgtgaagaaaggtgttctcaggtgaactaacttctcgggaccaagtcgca 1281
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 attgtgaataatgtgaagaaaggtgttctcaggtgaactaacttctcgggaccaagtcgca 1140

Qy 1282 aacaattcatccctccatcgtggccagagatatgtgcgggttgtccccagacatggcac 1341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 aacaattcatccctccatcgtggccagagatatgtgcgggttgtccccagacatggcac 1200

Qy 1342 cagaggatgctgaggtgggagctcaattggttgtccagagattacacaaggtaatgatca 1401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 cagaggatgctgaggtgggagctcaattggttgtccagagattacacaaggtaatgatca 1260

Qy 1402 ttggttgtcgccgaagacaagtcaaagcgaccagtcgtttcaacttaacgtggcct 1461
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 ttggttgtcgccgaagacaagtcaaagcgaccagtcgtttcaacttaacgtggcct 1320

Qy 1462 atcacaaggcccatccctcccccggaagaaacatccacagaggaataaaacattacacgtggct 1521
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 atcacaaggcccatccctcccccggaagaaacatccacagaggaataaaacattacacgtggct 1380

Qy 1522 attccattgtctcctctgttccttggctgtggaaatctttgctggagaaaacagactt 1581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 attccattgtctcctctgttccttggctgtggaaatctttgctggagaaaacagactt 1440

Qy 1582 agaaagaagaagaagaaaggaaatctccgtatggagaatatggatcacgcggagagttcagtct 1641
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 agaaagaagaagaagaaaggaaatctccgtatggagaatatggatcacgcggagagttcagtct 1500

Qy 1642 tggaagcaagatgataaaatatccctttgccagacacaactcagtcagtgagttaccatcactat 1701
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 tggaagcaagatgataaaatatccctttgccagacacaactcagtcagtgagttaccatcactat 1560

Qy 1702 gatcagaggactataatcctcacaagttagatcctcacacaagtgatgtgcaggttaa 1761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 gatcagaggactataatcctcacaagttagatcctcacacaagtgatgtgcaggttaa 1620

```
Matches 1522;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy  240 ggcggaggagctggggtcggtgtggtggacccagtgactattcaggaggatgagtggcacaat  299
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGCGGAGGAGCTGGGGTCGGTGTGGTGGACCCAGTGACTATTCAGGAGGATGAGTGGCACAAT   60

Qy  300 gacatctaagaattatcacccgggaccatcacaccacactgttcgcggaaaagacaattac  359
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GACACTCTAAGAATTATCACCCGGGACCATCACAGCACACTGTTCGCGGAAAAGACAATTTC  120

Qy  360 agtacacaaaggagaaaagactgattctgagtgttgggagattgatatcagaatcccagac  419
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AGTAACAAAAGGAGAAAAGACTGATTCTGAGTGTTGGGAGATTGGATATCACAATCCCAGAC  180

Qy  420 cctgtcttctgactacttctcaccagcttctcagaggatcaactttggcattatcagcaat  479
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGTCTTCTGACTATCTTCTCACCAGCTTCTCAGAGGATCAACATTTGGCATTATCAGCAAT  240

Qy  480 aagtagtgactgtcccacaacagtcttcgccggggttttcttgttgttgttctgaacaacaa  539
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAGTAGTGACTGTCCCACAACAGTCTTCGCCGGGGTTTTCTTGTTGTTGGAACAACAGCATC  300

Qy  540 gagtggatcccacacatttctgcccgcaaagactctggcttcaccagccatatggtccatactctg  599
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GAGTGGATCCCACACATTTCTGCCCGCAAAGACTCTTGGCTTCACCAGCCATATGGTCCATACC  360

Qy  600 agatttatggctttcactcttgttgttgaacacagagacagagaatacagctttga  659
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AGATTTATGGCTTTCACTCTTGTTGTTGAACACATTATTGGTCACAGAATACAGCAAATT  420

Qy  660 ctgcccagctggttgtagaacgtagcaaggaagacaattcctggggagaatatgggta  719
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 CTGCCCAAGCTGGTTGTAGAACGTAGCAAGGAAGACAATTCCTGGGGAATATGGTGCAATA  480

Qy  720 tagaatacactctttattgtgtcaggaaagactgccaatcatgcaggagaataattgctgaact  779
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 TAGAAGTACACTCTTTATTGTGTCAAAAGCTCCCATCCATGCAGGAGAATAATTCCTGATGAACT  540

Qy  780 aaggtggccagatcagtgtgcttcagcgcaaaggaatcagtcaatatgaaagggatctcggc  839
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 AGGTGGCCCAGATCAGTGTGCTTCAGCGCAAAGGAATCAGTCAGTATGAAAGGGATTCGGC  600

Qy  840 caatggtgctgttcttcgaggaggcgatcgatcttcttcttgttttctacctccaa  899
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 CAATGGTGCTGTTCTTCCGAGGGATGGTTCCCATCGACAGCAAGGCAAAGGGATTCCTCCTCATG  660

Qy  900 tggttgcagagatcctgagctttgcaggcaaaagagacttcagagccttctctccatg  959
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 TGGTTGCACAGATCCTGAGCTTTGCAGGCAAATCAGAGCCTTCTTCTACCTCCATG  720

Qy  960 gcagtcggtcaatgagagtggagaacaagtcaactgtctcctggccaagcgccgactca  1019
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GCAGTCGGTCAATGAGAGTGGAGAACAAGTTCAACTGTCTCCTGGCCAAGCCCGACTTCA  780

Qy 1020 ggacacaagcccatcatgggggcgacagtagcaacaaccaacaacaccgagagtg  1079
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GGACCAAGGCCCATCATGGGCTTCGGCGACAGTAGCAACAACCAACAACACCGAGAGTG  840

Qy 1080 gctggagatcgagattcgggggagaggaaaataaacaggaagaaatacagagggatctac  1139
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 GCTGGAGAGCGAGATCAGTTGGGGCAGAGAAAATAAACAGGAAGAAATACACACGAGGATCTAC  900

Qy 1140 acagtcgaactcaacttctatgttgatggaatctaagaaactcattcatcccctccagacaatt  1199
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 ACAGTCGAACTCAACTTTTATGTTGATGGAATCTAAGAAACTCATAATAACAATAAATTC  960

Qy 1200 taagtggaagacctataggaattgtgaataataatgaagaaaaagtgttttcaggggtaactc  1259
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 TAAGTGGAAGACCTATATAAAGAAT TAATTGAATAATTGTGTTTCACGGTAACTC  1020

Qy 1260 taacttctgggaccagtgcaaaacaattcatcctcccaccacgctggccagataatggtcatactactgt  1319
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 TAACTTCTGGGACCCAGTGCAAACAATTCATTCCTCCCACCACGCTGGCCAGATAATGGTCATACTGT  1080
```

```
Qy 1320 ggtgtgtccccccagacatggccacccagagacaggagagagtagcctctgaaggtggagctcattgttgtgca  1379
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GGTGTGTCCCCCCAGACATGGCCACCCAGAGACAGGATAGCCTCTGAAGGTGGACTCATTGTTGTGCCA  1140

Qy 1380 gattacacaaggtaatgaattcattggtcgtgtcggtcggcagccaagactgttc  1439
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 GATTACACAAGGTAATGATTCATTGGTCGTGTCGGTCGGCAGACAAGCTGTAAAAAGCACCAGTGTTTC  1200

Qy 1440 aactaagaaaagaaagatgagaacaatacaacagcaggcccctcccatcccctcgggaagaaaagacctccaagg  1499
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 AACTAAGAAAAGAAAGATGAGAACAATACAACAGCAGGCCCCCATCCCCTCGGGAAGAAAACATCCAAGG  1260

Qy 1500 aataacacacaacgtggtgctattccatgtggtcctctgttgtcctggttgtttctgg  1559
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 AATAAACATTACAACGGTGCTATTCCATGTGGTCCTCTGTGGTCCTGGTTGTTGTTGCCA  1320

Qy 1560 aatggggatcttgcagcctttgaagaaggaagaggaagagaaatttccgtatgatcagc  1619
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 AATGGGGATCTTTGCAGCCTTTGAAGAAGGAAGAGGAAGAGAAATTTCCGTATGATCAGC  1380

Qy 1620 ggaggctcagaaacagactgtgttggagagcagagattaaatatccctttgccaagaatcagtc  1679
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 GGAGGCTCAGAAAACAGACTGTGTTGGAGAGCAGAGATTAAAAATATCCCTTTGCCAAGAATCAGTC  1440

Qy 1680 agctgagtttaccatcaagtctatgtaataatgagcagagggaatagagatcatctcat  1739
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 AGCTGAGTTTACCATCAAGTCTATGTAATAATGAGCAGAGGGAATAGAGATCATCTCAT  1500

Qy 1740 cacaagtgatatggcaggttaa  1761
        ||||||||||||||||||||||
Db 1501 CACAAGTGATATGGCAGGTTAA  1522
```

RESULT 4
LOCUS   AX118816   1464 bp   DNA   linear   PAT 11-MAY-2001
DEFINITION Sequence 1 from Patent WO0129219.
ACCESSION   AX118816
VERSION   AX118816.1 GI:14035774
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1464)
  AUTHORS   Donoho,G., Turner,C.A., Nehls,M., Friedrich,G., Zambrowicz,B. and
            Sands,A.T.
  TITLE   Human cub-domain-containing protein and gene encoding the same
  JOURNAL   Patent: WO 0129219-A 1 26-APR-2001;
            Lexicon Genetics Incorporated (US)
FEATURES   Location/Qualifiers
     source   1..1464
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   440 a   303 c   359 g   362 t
ORIGIN

```
Query Match     83.1%;   Score 1464;   DB 6;   Length 1464;
Best Local Similarity   100.0%;   Pred. No. 1.3e-269;
Matches 1464;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy  298 atgacatctaaggaattatcccgggaccatcacaccacactgttcgcggaaaagacaatt  357
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGACAGTTAAGGAATTATCCCGGGACCATCACACCACACTGTTCGTTTCGGAAAAGAACAATT   60

Qy  358 acagtaccaaagggggaaaagactgattctgagtgttgggagattgatatcagaatcccag  417
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACAGTACCAAAGGGGGAAAAGTGATGATAATAATGAAGAAAAAGTGTTTTCAGGGTTAACTC  120

Qy  418 accgtgcttctgactatctcctccaccagatcaatatggttcaatatggctccatactactgt  477
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACACTCAACAAGGGAAAAGCTGATGTATTCATGTGGTCGGATATTCGAATCCCATCTCCAG  180
```

```
Qy   121  ACCTGTCTTCTGACTATCTTCTTCTTCACCAGCTCTTCAGATCAATATGGTCCATACTGTT  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  GGAACTATGACTGTTCCCAACAGTCTCTTTCCAAAGAACTCTTCGTTCGAACACAAGCTAACCGTCCCGCTTT  240

Qy   241  GAGAGCTGGATCCCAACATTCTCGCCCGGGGTCTTTTGTCGCTGACCATCGCAGCGAGCAGGCCAT  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  CCAGTTTAATAACATGTTTGGAACAGCAGAATACACAGGCCAT  360

Qy   361  TTCTGCCCATCGTGTAGAGACGTGTAGAGCAGGACATTTCTGGGAATAGGTAGAGTGAA  420
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   421  TATAGAGAATACCTTTATTGTCTCAAAGCTCCCATCCATGCGAGAATAATTGCTGATGAA  480

Qy   481  CTAGGTGGCCCAGATCAGTTGTGTGCTTCAGCGGCAAAGGATCCCAAAGGGATCCTG  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  GCCAATGGTCTGTCTTCTTCGAGGGATGGTCCCTGTTGACAGCAAGCCGGATTTCTGTTTACCTCC  600

Qy   601  AATGGTTGCAGACAGAACCTGAGTCTTTGCTGGGGGAGaaaaaaaaaaaaaatcaggaccttcttcctca  660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  TCGGCAGTTCGGTCATGAGTGGGTGAGCCACACCAGTGTCTCCTGGCCAGCCCGGACTT  720

Qy   721  CAGGACCAAGGCGGCCATCACGGGCTTCGGCGAGCCAACGTCAGTCTCATATGAAGGGAGATTCTG  780
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  TGGGTCGGACGATGGATTTGGGGGAGAAAAAAAGGTAAGGAAAAATAACAAGCCACACCAGGAATCT  840

Qy   841  ACACAGTCGGAACTCAACTTCACACTTTTAGTGGTTAAGAGTTTTGTAAGAGTTTTGGTGATGGTTC  900
          |||||||||||||||||||||||||||||||||||||
Db   901  TCTAGTGCAAGAACCTAATAAGAGATTCGAAATTAATGAGGAAAGAAGATAAAGGGTGTTCAGGGTAAC  960

Qy   961  TCTAACTTCGGGACCCAGTGCAAAACCAATTCCATCCTCCCCATCGCTGGCCAGACATGTG  1020
          |||||||||||||||||||||||||||||||||
Db   1021  CGGGTGTTCCCCAGGACCATGGCACCAGGGATAGCCTGCCCCAGATTGGGGCTCATTGGTTGC  1080

Qy   1081  CAGATTGACAAGACTATATAAGGATTAGGTGTGTGCCCAAGACAAGTACCACCAGTGTT  1140
          ||||||||||||||||||||||||||||||||||||||||||
Db   1141  TCAACTAAGAAAGAAGAAGTAGAGGACAAAAGCGATTGAAACAGAAAGTGAGTAAAGAAAGTGTC  1200

Qy   1201  GGAATAAAACATTACAACGGTTGCCTATTCCATGTGGTCCTTGTTCTCTCGGTGTTTGCT  1260
          ||||||||||||||||||||||||||||||
Db   1261  GGAATGGGGGATCTTTGCGCGCCTTAGAAAGAGAGAAGAAGAAGTCGTATGGATCA  1320
```

Right column:

```
Qy   1558  ggaatggggatctttgcgcgccttagaaagaggaaggaagaagtcgtatggatca  1617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  GGAATGGGGGATCTTTGCGCGCCTTAGAAAGAGAGAAGAAGAAGTCGTATGGGATCA  1320

Qy   1618  gcgggaggctcagagaaacagactgttggaagagattaaatatcccttgccagaatcag  1677
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  GCGGAGGCTCAGAAAACACACTGTTGGAAGCAGCGATTAAAATATCCCTTTGCCAGCATCAG  1380

Qy   1678  tcagctgagttcaccatcagctgtatgtagaagaaggaggaggaggtagatctc  1737
           ||||||||||||||||||||||||||||||||||||||
Db   1381  TCAGCTGAGTTCACCATCAGCTATGATAATGACGAGAAGGAGGAGGATGACAAAGGTTAGATCTC  1440

Qy   1738  atacacaagtgaatatggcaggttaa  1761
           ||||||||||||||||||||
Db   1441  ATCACAAGTGATATGGCAGTTAA  1464
```

```
REFERENCE   2  (bases 1 to 249429)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:17149292.
            -- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -- Project Information
            Center project name: GCOV
            Center clone name: CH230-8N22
            -- Summary Statistics
            Assembly program: Phrap; version 0.990329First call to
findPhrapList
            Consensus quality: 228631 bases at least Q40
            Consensus quality: 233904 bases at least Q30
            Consensus quality: 237865 bases at least Q20
            Estimated insert size: 235155; sum-of-contig estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 3.9x in Q20 bases; sum-of-contig estimation
            NOTE: Estimated insert size may differ from sequence length
                (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            NOTE: This is a 'working draft' sequence. It currently
            consists of 42 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *     1:  contig of 26784 bp in length
            *   26784:  gap of unknown length
            *   26885:  contig of 18137 bp in length
            *   45021:  gap of unknown length
            *   45022:  contig of 1628 bp in length
            *   45122:  gap of unknown length
            *   45123:  contig of 1624 bp in length
            *   61370:  gap of unknown length
            *   61371:  contig of 6398 bp in length
            *   61470:  gap of unknown length
            *   61471:  contig of 9692 bp in length
            *   71162:  gap of unknown length
            *   71163:  contig of 11383 bp in length
            *   82645:  gap of unknown length
            *   82646:  contig of 8274 bp in length
            *   82744:  gap of unknown length
            *   82745:  contig of 11868 bp in length
            *   94613:  gap of unknown length
            *   94614:  contig of 8192 bp in length
            *   94714:  gap of unknown length
            *  102905:  contig of 7909 bp in length
            *  103006:  gap of unknown length
            *  103006:  contig of 6648 bp in length
            *  109553:  gap of unknown length
            *  109554:  contig of 6273 bp in length
            *  109753:  gap of unknown length
            *  109754:  contig of 7219 bp in length
            *  116151:  gap of unknown length
            *  116152:  contig of 6398 bp in length
            *  116251:  gap of unknown length
            *  116252:  contig of 8986 bp in length
            *  125237:  gap of unknown length
            *  125238:  contig of 8886 bp in length
            *  125337:  gap of unknown length
            *  125338:  contig of 7909 bp in length
            *  133246:  gap of unknown length
            *  133247:  contig of 6099 bp in length
            *  133346:  gap of unknown length
            *  133347:  contig of 6273 bp in length
            *  139619:  gap of unknown length
            *  139620:  contig of 7219 bp in length
            *  139719:  gap of unknown length
            *  139720:  contig of 6044 bp in length
            *  146939:  gap of unknown length
            *  146939:  contig of 6547 bp in length
            *  147038:  gap of unknown length
            *  147039:  contig of 6044 bp in length
            *  153082:  gap of unknown length
            *  153083:  contig of 6547 bp in length
            *  159729:  gap of unknown length
            *  159730:  contig of 4122 bp in length
            *  159829:  gap of unknown length
            *  159830:  contig of 5657 bp in length
            *  165486:  gap of unknown length
            *  165487:  contig of 4122 bp in length
            *  169708:  gap of unknown length
            *  169709:  contig of 5420 bp in length
            *  169808:  gap of unknown length
            *  169809:  contig of 5420 bp in length
            *  175228:  gap of unknown length
            *  175229:  contig of 4947 bp in length
            *  175328:  gap of unknown length
            *  180275:  contig of 4947 bp in length
            *  180375:  gap of unknown length
```

```
            *  180376:  contig of 5150 bp in length
            *  185525:  gap of unknown length
            *  185526:  contig of 4557 bp in length
            *  190182:  gap of unknown length
            *  190183:  contig of 3821 bp in length
            *  190282:  gap of unknown length
            *  194103:  contig of 4198 bp in length
            *  194204:  gap of unknown length
            *  194401:  contig of 4198 bp in length
            *  198401:  gap of unknown length
            *  198501:  contig of 3595 bp in length
            *  202096:  gap of unknown length
            *  202097:  contig of 3345 bp in length
            *  205441:  gap of unknown length
            *  205542:  contig of 3090 bp in length
            *  208731:  gap of unknown length
            *  208732:  contig of 3505 bp in length
            *  208831:  gap of unknown length
            *  208931:  contig of 3505 bp in length
            *  212336:  gap of unknown length
            *  212337:  contig of 3007 bp in length
            *  215443:  gap of unknown length
            *  215444:  contig of 3592 bp in length
            *  215543:  gap of unknown length
            *  215544:  contig of 3592 bp in length
            *  219135:  gap of unknown length
            *  219136:  contig of 3062 bp in length
            *  222297:  gap of unknown length
            *  222298:  contig of 2391 bp in length
            *  224788:  gap of unknown length
            *  224888:  contig of 2438 bp in length
            *  227326:  gap of unknown length
            *  227327:  contig of 1859 bp in length
            *  229286:  gap of unknown length
            *  229385:  contig of 2341 bp in length
            *  231726:  gap of unknown length
            *  231727:  contig of 2695 bp in length
            *  234521:  gap of unknown length
            *  234522:  contig of 3045 bp in length
            *  237666:  gap of unknown length
            *  237667:  contig of 1804 bp in length
            *  239570:  gap of unknown length
            *  239571:  contig of 2779 bp in length
            *  239670:  gap of unknown length
            *  242449:  contig of 2038 bp in length
            *  242550:  gap of unknown length
            *  244549:  contig of 2038 bp in length
            *  244587:  gap of unknown length
            *  244588:  contig of 1041 bp in length
            *  245728:  gap of unknown length
            *  245729:  contig of 2433 bp in length
            *  248261:  gap of unknown length
            *  248262:  contig of 1068 bp in length
            *  248362:  gap of unknown length
            *  249429:  contig of 1068 bp in length.
FEATURES             Location/Qualifiers
     source          1..249429
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /chromosome="Mcs1"
                     /clone="CH230-8N22"
BASE COUNT    76470 a    47319 c   45841 g   75586 t     4213 others
ORIGIN
```

```
Query Match                     20.4%;   Score 359.6;  DB 2;  Length 249429;
Best Local Similarity    77.9%;   Pred: No. 1.5e-58;
Matches  500; Conservative   0; Mismatches  129;  Indels  13; Gaps  5;

Qy    270 ctagtgacttattcagatagtagtgcacaatgacatctaagaattatcccgggacctaccc 329
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   4548 CATAGTGACCTCTCAGGACATTGGGCACAATGACATCTAAGAATTATCCAGGGAGTCGTGGA 4489

Qy    330 caatcacactgttcg--cgaaaagacacattacacactaccaaaagggaaaaagactgctga 388
          |||||||||||||||  ||||||||| |||||||| ||| ||| |||||||||||||||||||
Db   4488 CAATTCACACTGTGTCGTGGGGAAGACGTCACACAGTCCCGAAAAGGGAGAAGAGACTTATTCTGA 4429

Qy    389 ggtgggagatcggaatcgcaatcgaatctgctcttctgactatctcttctcttcacca 448
          |||||||||||||||||||||||| |||| |||||||||||||||||||||||||||||
Db   4428 GGTT--GAGAATTGAAACACTGATTCCAAGACACTGTGCTTCCGCACTATCTCCTTCTTCAGCA 4371

Qy    449 gctctcaagatacaatatggtccactactgtggaagtatgactgttcccaaagaaactctgt 508
          ||||||||||||||| |||||||||||||||| ||| ||| |||||||||||||||||||||
Db                                                                      
```

```
Db   4370   GTCGCAACGGATCAGTAT-GACCCTATTGGTGGGAGTGTGGCGTCGTTTCCAGAGAACTCTGGC 4312
            ||||| || ||| || |||| ||||| || | |||| |||||||| ||||||| ||||||||
Qy    509   TGAACACAAGTGAAGTAACCGTCCCGCTTCGAGGAGTGGGATCCCACACATTCTGCGCCGGGTT 568
Db   4311   TGAACTCAAAGGAAGCAGGATCATCTCAGAGTGGATCTCACATTCCCTGCCCTGCCCGGGGCT 4252
            ||||| ||| | |||| ||| |||| | |||||| | |||||| | | |||| | |||| ||
Qy    569   TTTTGCTGACCCTATGCGAGCGAGCACACCCAGATTAATAACATGTTGGAACGAGCTA 628
Db   4251   TTCTGCCTCAGCCCTATTCTAGCAATGCATCCAGCATTTAATAACCTGTTTGGAATGAGGCA 4192
            ||||| ||||||| |||| | || || || ||||| | | || ||||| || || |||
Qy    629   gccattattttgaagacagaataacagcaaattctgcccactgtgttgtagaagctgacag 688
Db   4191   GTCATTAATTCCAAGGAAAAANATACAGGCCAGTCTCCCCACGCTGGTTGTGAGAGACAATAGCAA 4132
            ||||||| ||| ||| |||||||| | || | |||| || ||| | || | ||||| |||||
Qy    689   gagacatttctgggaatatggtgagatagaataacctctttattgtgcaaagctg 748
Db   4131   GGGACATTTCTGGGAATAACAAAAGACGGCTACTGAGATGGCCCTTATTCTGCCAAACCCG 4072
            ||||||||||| |||||| | || ||| |||||| | ||| || ||||| | ||| |||
Qy    749   ccatcccatgcaggaataattgctgatcgaactaggtggccagatca-------gtgtgct 800
Db   4071   ACATCCATGCAGGGGATCANTCTCANAATGANCTAGGTGGCTACCACACACCTCGTTCAGACTT 4012
            |||| || ||||| ||| | || | ||| ||||||||||| |||| || | | ||| ||
Qy    801   tcagcgcaaaggaatcaagtcgtatgaaggagggga 860
Db   4011   TCAGTGCAAAAGGGTTAAGTCACTCTGAGGAGTTCATGGTGTGTCGTCCACAACA 3952
            ||| |||||| |||| ||| | | || |||||||||| | ||| | ||| ||||
Qy    861   tggttccctgtcagacaagcgattctcgttacctccaatgg 902
Db   3951   TGGTTCCTGTGTCAGA--AAGTCAATTCTGTTTACATCCCCAGG 3911
            |||||||| | |||||    |||| || ||| || ||| || |
```

FEATURES   Location/Qualifiers

RESULT 6
LOCUS   HSDJ92C8
HSDJ92C8/c
DEFINITION Human DNA sequence from clone RPI-92C8 on chromosome 6, complete
            sequence.
ACCESSION AL132671
VERSION   AL132671.20   GI:9885189
KEYWORDS   HTG.
SOURCE   human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 131060)
AUTHORS   Tromans,A.
TITLE   Direct Submission
JOURNAL   Submitted (21-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
           CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
           requests: clonerequest@sanger.ac.uk
COMMENT   On Aug 22, 2000 this sequence version replaced gi:7228316.
           During sequence assembly data is compared from overlapping clones.
           Where differences are found these are annotated as variations
           together with a note of the overlapping clone name. Note that the
           variation annotation may not be found in the sequence submission
           corresponding to the overlapping clone, as we submit sequences with
           only a small overlap as described above.
           This sequence has been finished according to sequence map criteria
           as follows. An attempt is made to resolve all sequencing problems,
           such as compressions and repeats, but not necessarily within known
           annotated human repeat sequence elements (e.g. Alu). Where the
           sequence is ambiguous, there is an annotation using the 'unsure'
           feature key.
           The following abbreviations are used to associate primary accession
           numbers given in the feature table with their source databases:
           Em:, EMBL; Sv:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
           on the WORMPEP database can be found at
           http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
           was generated from part of bacterial clone contigs of human
           chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
           Group. Further information can be found at
           http://www.sanger.ac.uk/HGP/Chr6
           RPI-92C8 is from the library RPCI-1 constructed at the Roswell Park
           Cancer Institute by the group of Pieter de Jong. For further

           details see http://bacpac.med.buffalo.edu/
           VECTOR: pCYPAC2
           IMPORTANT: This sequence is not the entire insert of clone RPI-92C8
           It may be shorter because we sequence overlapping sections only
           once, except for a 100 base overlap.
           The true left end of clone RPI-179P9 is at 130955 in this sequence.
           The true right end of clone RPI-94G16 is at 100 in this sequence.

source   1..131060
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /chromosome="6"
           /clone_lib="RPCI-1"
           /clone="RPI-92C8"
misc_feature 1..281
           /note="CpG island"
           /evidence=not_experimental
repeat_region 4614..4755
           /note="71 copies 2 mer cc 59% conserved"
repeat_region 4621..4707
           /note="Alub repeat: matches 11..293 of consensus"
repeat_region 425..513
           /note="3 copies 29 mer 81% conserved"
repeat_region 6477..6564
           /note="MER57-internal repeat: matches 1252..1347 of
           consensus"
repeat_region 514..856
           /note="2 copies 44 mer 95% conserved"
repeat_region 7144..7280
           /note="MER1B repeat: matches 22..163 of consensus"
repeat_region 857..1162
           /note="FAM repeat: matches 2005..2220 of
           consensus"
repeat_region 7828..7888
           /note="MSTA repeat: matches 1..61 of consensus"
repeat_region 1977..2048
           /note="L2 repeat: matches 2246..2562 of consensus"
repeat_region 8038..8120
           /note="L2 repeat: matches 2465..2551 of consensus"
repeat_region 3004..3273
           /note="L2 repeat: matches 2629..2704 of consensus"
repeat_region 8223..8531
           /note="AluSq repeat: matches 1..305 of consensus"
repeat_region 4127..5300
           /note="Alub repeat: matches 18..287 of consensus"
repeat_region 8639..9030
           /note="MER41-internal repeat: matches 3972..4590 of
           consensus"
repeat_region 9455..9559
           /note="MER31A repeat: matches 1..417 of consensus"
repeat_region 9543..9743
           /note="MER57-internal repeat: matches 2005..2220 of
           consensus"
repeat_region 9732..9907
           /note="MER57-internal repeat: matches 2456..2629 of
           consensus"
repeat_region 9900..10720
           /note="MER57-internal repeat: matches 2777..3637 of
           consensus"
repeat_region 10709..11327
           /note="MER41-internal repeat: matches 4553..5213 of
           consensus"
repeat_region 11328..11618
           /note="AluSq repeat: matches 1..292 of consensus"
repeat_region 11619..12914
           /note="MER41-internal repeat: matches 4590..4554 of
           consensus"
repeat_region 12915..13144
           /note="L1PA5 repeat: matches 5914..6143 of consensus"
repeat_region 13145..13805
           /note="MER41-internal repeat: matches 4553..5213 of
           consensus"
repeat_region 13824..14458
           /note="MER41B repeat: matches 1..633 of consensus"
repeat_region 15991..16301
```

```
repeat_region   18848..19305    /note="AluY repeat: matches 1..311 of consensus"
repeat_region   19335..19575    /note="LiPB1 repeat: matches 5685..6152 of consensus"
repeat_region   19570..19759    /note="MER21A repeat: matches 684..933 of consensus"
repeat_region   20528..20618    /note="MER21B repeat: matches 326..531 of consensus"
repeat_region   20926..21059    /note="MIR repeat: matches 48..146 of consensus"
repeat_region   21601..21760    /note="LiMC1 repeat: matches 6189..6327 of consensus"
repeat_region   22605..22747    /note="AluJ repeat: matches 144..304 of consensus"
repeat_region   22956..23435    /note="MER91A repeat: matches 28..159 of consensus"
repeat_region   23463..23624    /note="LiMD repeat: matches -1..475 of consensus"
repeat_region   23626..23692    /note="LiM4 repeat: matches 860..1022 of consensus"
repeat_region   23693..24058    /note="MER7A repeat: matches 283..346 of consensus"
repeat_region   24059..24264    /note="THE1C repeat: matches 1..369 of consensus"
repeat_region   24547..24602    /note="MER7A repeat: matches 63..283 of consensus"
repeat_region   24622..25110    /note="28 copies 2 mer aa 71% conserved"
repeat_region   25129..26287    /note="LiM4 repeat: matches 2209..2702 of consensus"
repeat_region   26288..26589    /note="LiM4 repeat: matches 3174..4259 of consensus"
repeat_region   26590..26692    /note="AluSq repeat: matches 1..302 of consensus"
repeat_region   26694..26816    /note="LiM4 repeat: matches 3067..3174 of consensus"
repeat_region   26862..28116    /note="LiM4 repeat: matches 2707..2830 of consensus"
repeat_region   28118..28233    /note="LiMD1 repeat: matches 4968..6213 of consensus"
repeat_region   28958..29218    /note="4 copies 29 mer 73% conserved"
repeat_region   29294..29443    /note="LiME3A repeat: matches 5319..5594 of consensus"
repeat_region   2979..2990      /note="LiME3A repeat: matches 5744..5894 of consensus"
repeat_region   3284..32990     /note="U6 repeat: matches 1..107 of consensus"
repeat_region   3301..33293     /note="MIR repeat: matches 74..230 of consensus"
repeat_region   3294..33430     /note="AluO repeat: matches 3..305 of consensus"
repeat_region   3381..33949     /note="Alu repeat: matches 1..138 of consensus"
repeat_region   3437..33618     /note="MIR repeat: matches 67..142 of consensus"
repeat_region   3753..35058     /note="AluO repeat: matches 2..283 of consensus"
repeat_region   37633..37939    /note="AluO repeat: matches 2..296 of consensus"
repeat_region   38218..38311    /note="AluO repeat: matches 3..309 of consensus"
repeat_region   38453..38920    /note="MER1 repeat: matches 1..104 of consensus"
repeat_region   39205..39278    /note="LiM87 repeat: matches 5694..6171 of consensus"
repeat_region   3919..40220     /note="MIR repeat: matches 117..193 of consensus"
repeat_region   41119..41144    /note="Alub repeat: matches 3..302 of consensus"
repeat_region   41234..41534    /note="13 copies 2 mer ac 100% conserved"
repeat_region   41920..42076    /note="AluO repeat: matches 1..298 of consensus"
repeat_region   42077..42367    /note="L2 repeat: matches 2187..2344 of consensus"
repeat_region   42368..42399    /note="AluY repeat: matches 12..297 of consensus"
repeat_region   44012..44103    /note="L2 repeat: matches 2156..2187 of consensus"
repeat_region   44551..44785    /note="MER33 repeat: matches 1..92 of consensus"
repeat_region   44928..45276    /note="MIR repeat: matches 2..249 of consensus"
repeat_region   46018..46250    /note="THE1B repeat: matches 1..364 of consensus"
repeat_region   ...             /note="L2 repeat: matches 2498..2742 of consensus"
```

Query Match   14.6%;   Score 257; DB 9;   Length 131060;
Best Local Similarity 100.0%; Pred. No. 5.5e-39;
Matches 257; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1  atgggattcggttcggggcagcgactgcgccccgtccggcgcgcccggctccgcagag  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4994  atcggattcggttcgcggggcagcgactgcgccccgtccggcgcgcccggctccgcagag  4935

Qy     61  gaggcggcgcccggccccggagcctgcgctcggggatccgtcggagggggaggccgaagcttgcc  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4934  gaggcggcgcccggccccggcagcctgcgctcggggatccgtcggaggggagggccgaagcttgcc  4875

Qy    121  aagctggcgccaggcgggtcatgtcggccagggccgcggcggcggcggcggcactggcgcgg  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4874  aagctggcgccaggcgggtcatgtcggccagggccgcggcggcggcggcggcactggcgcgg  4815

Qy    181  gctgcgcgggcggggccctcactggcttcgcgtctccggcgcccccgtccggctgcag  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4814  gctgcgcgggcggggccctcactggcttcgcgtctccggcgcccccgtccggctgcag  4755

Qy    241  gcggaggaggctgggtga  257
           ||||||||||||||||||
Db   4754  gcggaggaggctgggtga  4738
```

RESULT 7
AF387548
LOCUS   AF387548   2642 bp   mRNA   linear   ROD 11-NOV-2001
DEFINITION   Mus musculus strain ICR endothelial and smooth muscle cell-derived
              neuropilin-like protein (Esdn) mRNA, complete cds.
ACCESSION   AF387548
VERSION   AF387548.1 GI:16902436
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2642)
  AUTHORS   Kobuke,K., Furukawa,Y., Sugai,M., Taniga ki,K., Ohashi,N.,
              Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.
  TITLE   ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
              Cells with the Longest Secretory Signal Sequence among Eukaryotes,
              Is Up-regulated after Vascular Injury
  JOURNAL   J. Biol. Chem. 276 (36), 34105-34114 (2001)
  PUBMED   11147234
REFERENCE   2 (bases 1 to 2642)
  AUTHORS   Kobuke,K., Furukawa,Y., Sugai,M., Taniga ki,K., Ohashi,N.,
              Sasayama,S., Matsumori,A., Honjo,T. and Tashiro,K.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto
              University Graduate School of Medicine, Yoshida Konoe-cho,
              Sakyo-ku, Kyoto 606-8501, Japan
FEATURES   Location/Qualifiers
     source   1..2642
                     /organism="Mus musculus"
                     /strain="ICR"
                     /db_xref="taxon:10090"
     gene   1..2642

```
CDS             /gene="Esdn"
                /note="similar to the Mus musculus strain 129/SvJ Esdn
                gene sequence in GenBank Accession Number AF442820"
        314..2623
                /gene="Esdn"
                /codon_start=1
                /product="endothelial and smooth muscle cell-derived
                neuropilin-like protein"
                /protein_id="AAL30179.1"
                /db_xref="GI:16902437"
                /translation="MASRAPLRAANSPGPGCGPAPAPAATGRAALPSAGCCPLPPGRNS
                SSPRLILLLLLLQDAGSGGDDGCHLANYLKIFNGIGVSRTEIGYXCGLGLQMNGISRSKG
                RTGERLRIRKFGDFDLEDSDYCHLANYLKIFNGIGVSRTEIGYXCGLGLQMNGISRSKG
                EVTYLPMSCTHAARCGPLASYSVIDKEDLTPCLDTYSNFLEPFERSYCPAGCLLPFAE
                ISGTIPHGVRDSSPPLCMAGTIHGVSNVLGQGISIVISKGPYTYESSLANNVTSVGI
                LSASLFTPRKTSCGYCTLGMEESGVTIAQPGITASSALEWTDHWGGQNSWTAEKARLRKPG
                RAGWMKGMQLLPAKSVDHHEEPYRVYSIEVSHLSARENVTVLQADSAEYAQVWI
                REPGVDGQDKIFCGNKHDHKUVRNNFLPPIIARPIRVNPVQMOOKIAMKVELLGCQFTL
                VGTLHQRSTFKPEGEKGTAGYLDDYNSPMQEVYHAYPAEPLPVTGEYATIPIVMDMSG
                IKNTTVTPSVTKDVALAAVIVPVLVMALTTLILILVCAMHHRNRKNRKKTIDGAYDLPHND
                HPTASVGLPSTISTFKTAGTQPHALVGTYNTLLSRDSCSSGQAQYDTPKGGKSAATPE
                ELVTVQPQSTQELSGAGRDEKFDAPKEILL"
misc_feature   431..778
                /gene="Esdn"
                /note="corresponds to nt 453 to 800, referred to in J.
                Biol. Chem. 276 (36), 34105-34114 (2001)"

BASE COUNT      639 a    703 c    744 g    556 t
ORIGIN
```

```
Best Local Similarity   50.9%;   Pred. No. 5.9e-34;
Query Match   13.0%;  Score 229.8;  DB 10;  Length 2642;
Matches 657;   Conservative   0;   Mismatches 592;   Indels   42;   Gaps   3;
```

```
AK056350
AK056350
LOCUS       AK056350                2020 bp    mRNA            linear   PRI 31-OCT-2001
DEFINITION  Homo sapiens cDNA FLJ31788 fis, clone NT2RI2008598, weakly similar
            to MILK FAT GLOBULE-EGF FACTOR 8 PRECURSOR.
ACCESSION   AK056350
VERSION     AK056350.1  GI:16551729
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RI clone:NT2RI2008598.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Isshiki,T., Kanehori,K., Yoshida,M., Watanabe,S., Ishida,S.,
            Ono,Y., Hotuta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,
            Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,
            Yamashita,H., Chiba,J., Sugiyama,T., Irie,R., Otsuki,T.,
            Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,
            Saito,M., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,
            Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Takahashi-Fujii,A.,
```

REFERENCE
AUTHORS   Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
           Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 2020)
AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE   Direct Submission
JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
           Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)

COMMENT   NEDO human cDNA sequencing project supported by Ministry of
           Economy, Trade and Industry of Japan; cDNA full insert sequencing:
           Research Association for Biotechnology (RAB) (supported by Japan
           Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
           HRI, and Biotechnology Center, National Institute of Technology and
           Evaluation; clone selection for full insert sequencing: RAB and
           HRI.

FEATURES   Location/Qualifiers
     source   1..2020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2R12008598"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2R12"
                     /note="cloning vector: pME18SFL3~mRNA from NT2 neuronal
                     precursor cells treated 2-weeks retinoic acid (RA) induction.~majorly NT2 neuron"

BASE COUNT   494 a   519 c   530 g   477 t
ORIGIN

Query Match   12.8%;   Score 225.8;   DB 9;   Length 2020;
Best Local Similarity   52.1%;   Pred. No. 3.3e-33;
Matches 608;   Conservative   0;   Mismatches 517;   Indels 42;   Gaps 3;

Qy   248 agctggggtgatcggctggcaccccagtagacttatcaaggatggtcgcacaatgacatca 307
          |||| ||||||||| || | || | ||| | || ||||| | | || ||||||| |
Db   556 AGCAGAGGTGATGGACACTACTAGGCCCTGAGAGACTGGAAAGCCCTTACATCCA 615

Qy   308 agaattacctcccgggacctacccccaatcacactgttcgcgaaaagacaattacagtaccaa 367
          | |||||| ||| | |||| | | ||| |||| ||| ||| ||| |||| | ||| |
Db   616 TAAACTACCCACAGACACTATCCCAACAGACAGCTGTTTGTGAATGGAGATCCGGTGTAAAGA 675

Qy   368 aggggaaaaagactgattcgagggttgggaggatttggatcgaa------------ 411
          |||| | | ||| | ||| || |||| | |||||| ||| ,411
Db   676 TGGGGACAGAGAGTTCCGATCAAATTGGTGACTTGACAATTGAAGAATTCTGATTCTGTC 735

Qy   412 --tcccagacctgtgctgtcttgactatctctcttccacagctcctcagaataacataggtc 469
            ||| ||| | ||| | || | |||||| | | ||||||| | | | || | ||||
Db   736 ACTTYAATTATTACTTGACAATTTAAATGGAATTGGAGAATCGAAACTGAAATAGGGTC 795

Qy   470 catactgtggaagtatgactgttcccaaagaacctctgttgtgaaccacaagtgaagtaaccg 529
          || ||||||| | ||||| | ||| || || | || || |||| |||| ||| ||||
Db   796 ACTGTGGTCTGGGTGCACAAATGAACCANTCAAATTGAATCAAAGGCAATGAAAATACACAT 855

Qy   530 tccgcttgagagtcgggatcccaacatttctggccccgggggtttttctgtcgacctatgcgagca 589
          | ||||| | ||||||| |||||||||| | |||| ||| ||||| | |||||| ||||
Db   856 TGCTGTTTCATGAGTGGAANTCCATGTTWTCTGAGCGCGGGGGATNTTTTWGGCCTCATTACTCGTGTTTA 915

Qy   590 gcgacacaccagattttaataacatgttgaacaagctagcacatattattttgaaacagaat 649
          | |||| ||| | ||| ||| | ||||| ||||| ||||||||||||||||| ||| ||||
Db   916 TACATAAAAACAAGATACTAATTACTGTGTTTGGCACAGCTACATATTATTTTTTGGAACCTGAGT 975

Qy   650 acagcaaaattctgcccccacctcgttgtagagacgtagcaggagaaacattcttgggatatgg 709
          ||||| | ||||| ||| | |||||| | || | |||||| | | ||| || |||||
Db   976 TCAGTAAGTACTACTCCCGACATGTTGTCTGCTCCGATNTCGACGAGATNTCGGAACAATTC 1035

Qy   710 tagatggatatagaagatacctcttttattgtgcaaagctggccatccatgcaggagaataattc 769
          | |||||| |||| ||| ||||| || ||||| | || ||||||||| ||||||| | ||
Db   1036 CTCCATGGATAATATAGGATTCCTCGCGATNCCCATTGGTNGCATCGCAGGAGCTAGTCTT 1095

RESULT   9
AF387547
LOCUS   AF387547   2939 bp   mRNA   linear   PRI 11-NOV-2001
DEFINITION Homo sapiens endothelial and smooth muscle cell-derived
            neuropilin-like protein (ESDN) mRNA, complete cds.
ACCESSION   AF387547
VERSION   AF387547.1 GI:16902434
KEYWORDS   .
SOURCE   Homo sapiens (human).
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2939)
AUTHORS   Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
            Matsumori,A., Sasayama,S., Honjo,T. and Tashiro,K.
TITLE   ESDN, a Novel Neuropilin-like Membrane Protein Cloned from Vascular
            Cells with the Longest Secretory Signal Sequence among Eukaryotes,
            Is Up-regulated after Vascular Injury
JOURNAL   J. Biol. Chem. 276 (36), 34105-34114 (2001)
PUBMED   11447234
REFERENCE   2 (bases 1 to 2939)
AUTHORS   Kobuke,K., Furukawa,Y., Sugai,M., Tanigaki,K., Ohashi,N.,
            Sasayama,S., Matsumori,A., Honjo,T. and Tashiro,K.
TITLE   Direct Submission
JOURNAL   Submitted (30-MAY-2001) Department of Medical Chemistry, Kyoto
            University Graduate School of Medicine, Yoshida Konoe-cho,
            Sakyo-ku, Kyoto 606-8501, Japan
FEATURES   Location/Qualifiers

Qy   770 ctggtgaactaggtggccagatcaagtgtgcttcagcgccaaaggggatcaagtcagtatgagg 829
          | | || || | || ||| | | || | ||| ||||| |||| ||||| | || || |
Db   1096 CAAACACGTTGGGCGCGGCCCAAAATCAGTCGTTGTAAGAAGTACCCCTATTATAAA 1155

Qy   830 ggattctggccaatgtgttcttccgagatggtggttccctgtcaggacaagcgattctgt 889
          | |||||||| ||||||||| ||| ||||||||||| || |||||||||| |||||| |
Db   1156 GTYCTTGGCTAACAACCTCACATCTGTGGTGGGACACTATTCNACAGTCTTTTACAT 1215

Qy   890 tcacctccaatgtgttcagcagtcctgagtttgaacctctacagcc-------c 937
          |||| |||||||| || |||| ||| ||| |||| | |||||| | c
Db   1216 TTMACAGANGATGGATTMGGACACTGCGGAGNTCTGGTGTCNNCGGATCCTC 1275

Qy   938 aaatcagaagcttcctcctcatggcagtcggtcggtcaatgagagtcgaagacaagttcactcgt 997
          |||| ||| | | |||| | |||||| | || | ||||||||| | | | |||| |||| |
Db   1276 AAATAACAGCATCATCTGTCGTCGGAGTGGGACTGACCACACACGACGGGAACNAGTTGA 1335

Qy   998 ctctggccaagcccgacttcaggacaagacggccatcatgggctcttcgggcacagtagca 1057
          || |||||||||| ||| | ||| |||| ||||| | |||||| | ||||| |||| |
Db   1336 AACCCAAAAAAAACCCAGCCTGAAAAAANCCTGGACCCGCGGGGCTTGCTTTNCGCTNGATG 1395

Qy   1058 acaaccaccacaaccaccgagtggctgaaatggacgatcatggggggaagaaaaataacag 1117
          | ||||| | | ||| |||||| |||| ||| | | || |||||||||| | |
Db   1396 AATACC----------ACTGGTMACNAATAGANTTTGAATAAGGAAAGAAAAANACAG 1443

Qy   1118 gaattaggaccacagggatctacacagtcgaactcaacttttatgttaagagtttgtga 1177
          | |||| |||| || | |||| ||| | |||| | |||||| |||| |||||||||| |
Db   1444 GCATTATAACCTACGTATNCACCTGGATNCGAGCACNAATTACCACCTATGTGTCTCCTTACAGAA 1503

Qy   1178 tgaacttcaaaaacaaataattctaagtggaagacctataaaggaattgtgaataatggg 1237
          ||| |||| |||| |||| | ||| |||||||||| |||||||||||||||||| ||||
Db   1504 TCCTGTCGCAGTGATGATGGGCAGAAAAATGGCACTGTGTACNGAGAGCCCTTGGTTCACGAT 1563

Qy   1238 aaaagtgttttcagggtaacctcaacttctgggaccaggtgcaaaacaaacattcatccctc 1297
          |||| |||||||| ||| ||| ||||||||||||||| ||||||| |||| |||| | ||
Db   1564 ATAGAGNTATTTTCANGGAANACAANAGATTAATCACCAAGGATNGCGTAATAACTTTTTGCCAC 1623

Qy   1298 ccatcgttggccagatatgtcgcgggttgtcgtcgacctatgcgagca 1357
          |||| ||||||||||||||| | || ||| | | |||||||||||||
Db   1624 CAATTATTTGCACGTTYTNATTNAGNAGTGNAATCCTACCACACNAGCACAGNAGAAANATGCCATGA 1683

Qy   1358 aggtgaggtcattgttgccagaata 1384
          ||||||||| | |||||||||| ||
Db   1684 AAATGGAGCTGCTTCCGCGATGTCAGTTTA 1710

FEATURES
   source

source
   1. .2939
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="3"

gene
   1. .2939
   /gene="ESDN"

CDS
   12. .2339
   /gene="ESDN"
   /codon_start=1
   /product="endothelial and smooth muscle cell-derived
   neuropilin-like protein"
   /protein_id="AAL30178.1"
   /db_xref="GI:16902435"
   /translation="MASRAVRARRCPQCPQVRAAAAAAPAWAALPLSRSLPPCSNSSS
FSMPLLFLLLVLLLLLEDAGAQQGDGDCGGHTVLGPESGTLTISINVPQTYPNSTVCEWE
IRVKMGEVRVRIKFDDFDIEDSDSCHFNYLRITNGIGSVRTEIGKYCGLGLQQMNHISIES
NEINILTMPMSGIVWSRGFRLASYSVIDKQDLITICLDTASNFLPEPEFSKYCPAGCLLP
FAEISGTIPHGYRSDSSPLCMAGVHAGVVSNTLCGGISVVISKGITFYYESSLANNVTSV
VGHLSILTLPTFRKTISGCYIGLGMESGVIADPDTTASSVLEWTDWTDTAKGQENSWRPKARLK
KPGPWAAFATDEYQWLQIDLNEKLIGIITTGSTGWMVEVNYVVSAYRILYSDDGQKW
TASPDIRNTLVTPNVTKDNVALAVAPVLVRLNVLLTTLLLILVCAWHRNRNKKKTTEGTYD
FIPKGRPPLTQPPRPENSNDLKNTLTAPKIKAKRAPKFTQPLQPRSSNEFPAQTKGYD
LVGGIVGTLLQRSTFKPEGKEAGYTADLLDPYNPGQEVVHAYAEPLPTIGPEYATFII
MDMSGHFTTSVGQPSFKATGNQPPLVGTYNTLSRTDSCSSAAQAYDTPRAGKP
GLPAPDELVYQVPQSTOEVSGAGRDGECDVFKEIL"

ORIGIN

BASE COUNT   825 a   675 c   657 g   782 t

Query Match   12.7%;
Best Local Similarity 52.0%;
Matches 607; Conservative 0;

Score 224.2; DB 9; Length 2339;
Pred. No. 7e-33;
Mismatches 518; Indels 42; Gaps 3;

Qy   248 agtcgtgggtttgtgctggtctggacacacactgagtgagtcagcacatcgacatcta
    || || |||||||||||| |||||||||||||||||||||||||||||
Db   751 CAAACACGTTGCGCGCGCGCCAAATCAGTCGTGTGTAATAGTAGTAAAGGTATTCCCTATTATGAAA 810

Qy   248 agtcgtgggtttgtgctggtctggacacacactgagtgagtcagcacatcgacatcta 307
    || ||| ||||||||||||||
Db   830 ggattctggccaatggtggtgttcctctcgaggagtgttccctgtcagacaagcggattctgt 889

Qy   308 agaaattacccggaccaccccaataccacactgttggaaaagacaattacagtacaa 367
    |||||| |||||||||||
Db   811 GTTCTTTGGCTAACAACGTTCACATCTCGTGTGGCAGACTTACTTAATCTACAATCCA 870

Qy   368 aggggaaaaagactgatctgaggtggggaagattcggatcaa-------- 411
    ||||||||| || || |||||
Db   890 ttacctccaatggttgtgcaggcagatccttcgagtctttgaacctgacgg------------c 937

Qy   411 TGGGAGAGAGAGTCCGCCATCAAATTTGGTGACTTGCAATTGCAATTGATTTCGTTGTC 331
    |||||||||||||
Db   871 TTAAGCAACAAGTGGATGTTTATGCAACACACTGGGACAGGTCTTGGTTGGTGGATCCTC 930

Qy   412 --tcccagacctgtgctctgactatctctccacagctcttcagatcaatatggtc 469
    ||||||||||||
Db   938 aaatcaggagctcttcctcatggcaaatgatttcatgacaagtggagagtggaga 997

Qy   391 ACTTTAATTAATTACTTGAGAATTTATAPAATGGACACTGGAACCCTTACATCCA 450
    |||||||||||||
Db   931 AAATAACAGCATAAATTCTCTGTCTGGTTGACATTGCAATATCTCGGACAACATTC 990

Qy   470 catactgtggaagtatgactgttcccaagaacctgttgaacacaagtgaagtaaccg 529
    ||||||||||||
Db   998 ctcctggcaagcgccgactcaggacaccaaggccccaatcatgggtctcggagcagtagca 1057

Qy   451 ACTCGTGTCTGTGGGGTTGCAAATGACAATTCAATTGAATCAAAAAGGCAATGAACACAT 510
    ||||||||||||
Db   991 AACCCAAAAAAAGCCAAGCAGTGCAAAAAAACCTGGAACCGCGCCCTTGGCGTGCTTTGTACGCATC 1050

Qy   530 tccgcttgagtgtgatccccacattctggccggggttttttgctgactatgcgagca 589
    |||||||||||||
Db   1058 acaaccacaaaccaacgagagtggctgggagacgcattctggggaagaaagaataacag 1117

Qy   511 TGCTGTTCATGAGTGGAATCCAACGAGTTCTGGACGACGCCGAATTTTGGCCTATACTCTGTTA 570
    |||||||||||||
Db   1051 AATACC--------AGTCGTTACAAATAGATTTGAATAAGGAAAAGAAAAGAAAAGTCAGA 1098

Qy   571 TAGATAAACAAGATCTAATTACTCTGTTTGGACACTGCCATCCAATATTTTTTGGACCGTAGCT 630
    ||||||
Db   1099 GCAATTAAAACACCACTGCGATCCACCATGGTGGCAGACAATTACTATCTGTTCGGCACAGGA 1158

Qy   650 acagcaaaatctgccaagtcggttgtgtagagacgtagcaggagacattctggggaatatggt 709
    |||||||||||||
Db   1178 tgaactcaaaaacataatctaagtgaagacctataagggaatgtgaatatgaag 1237

Qy   630 TCAGTAAGTTACTGCCCAGTCGGTTCGTCTCGTCCTTCCTTTTGCTGCAGAATATCTCGGACAACAATTC 690
    ||||||||||||
Db   1159 TCCTGCTACAGTGTCAGTGATGATGGCAGACAGAAATGGCAGAAGCCTGTGGTTGGAGCAAG 1218

Qy   710 tagatgagattatagagataccttcatttatttgtgaaaggtgccatcccaatgcaggaataatctg 769
    |||||||||||||
Db   1219 AATAAGATAATTTCAAGGAAACAAAAGATTATTATCACCACAGGGAAGTGGACGTGCCATCTCCAC 1278

Qy   691 CTCACAGGAACTAATAGATTAGCAATTCCTGGCCCATTCTGCGCCATGAATGTGCGCCATGTGTCAGGAGATCGAGTGT 750
    |||||||||||||
Db   1298 ccatcggtggccagagatatgtgcggggttgtcccacacgacatgtgcgacctttta 1357

Qy   770 ctgatgagactaggtctgtggccagatcagttgcttcagccaagaggatcagtcgatatgaag 829
    || ||| || |||||
Db   1279 AACCCAAAAAAAGCCAAGCAGTGCAAAAAAACCTGGAACCGCGCCCTTGGCGTGCTTTGTACGCATC 1338

Db   1358 aggtggagagtcattggtgccagatta 1384
    |||||||||||||
Db   1339 AAATGGAACTGCTCGGATGTCAGTTTA 1365

gene   1..2339
            /gene="Esdn"
CDS   12..2321
            /gene="Esdn"
            /codon_start=1
            /product="endothelial and smooth muscle cell-derived
            neuropilin-like protein"
            /protein_id="AAL30180.1"
            /db_xref="GI:16902439"
            /translation="MASRAPLRRARSPQDPRGRAAPANRGRAPLPSAGWCPLPGCRNS
            SRPRLILLLLLLPDAQAQNGGCHVLGPESGFTGISINPHFPNSTVCKWEITRV
            KTGERKIRIKFGDFDIEDSDYCHLNYLKIFNGIGVSREIERIGKYCGSLQNMVCDIESKGS
            EITVLFMSGHIHASGRFLASYSVIDKQDLICLDTVSNFLEPESKYCPAGLLPFAE
            ISGTLPHGYRDSSPLCMAGIHAGVVSGDGISVVISKGPYYIYSSLANNVSNVGI
            LSTSLFTFKYSGCIYGTLGMESGVIAPDQTASSVLEWTDIHMGQENSMKPEKARLIKPG
            PPWAAPATDHOWLQIDLNKERKITGIVTTGSTLIEHNYVSAYRVLVSDDGQKWTVY
            REPGAADQKIFQGNDYHKDVRNNFLPLIARFIRVNPVQWQKAMKVELLGCGFTL
            KGRLPLKLTQPPPRNSNLNKNTVHPKLGRAPKFTQALQPSRNDLPLPAHPNTATQ
            RAGMWKAVKGQLLPAKSVEHEETEVPKRSNSEVSHLSPREVTTVLQADSEAYAQPLVGGI
            VGTLHORSTFKPPEGKEASTADLDPYINAPVQEYVHAYAEPLPVTGPEYATPIVMDNSG
            HSTASVGLPSTSTFRTAGNQPPALVGTVNTLLSRTDSCSSGOAQYIDTPKGGKPAAAPE
            ELVYQVPGSTQEASGAGRDEKFDAFKETL"

Query Match   12.5%;   Score 220.4;   DB 10;   Length 2339;
Best Local Similarity 50.1%;   Pred. No. 3.6e-32;
Matches 688; Conservative   0;   Mismatches 641;   Indels 43;   Gaps 4;

Qy   46 cgctcgtctcccccaaggaggcggcccgggccccgggcctcgcgcctcgcggatccgcgagg 105
          ||||||||| | |||| | || || | ||| ||| ||||| | | | | | |
Db   1 CGGCCGCCGGGCGGAGCGGGGCGGGCCGGGCCCCGGGCCCGGGGGCGGCCCCAGGAYCCG 60

Qy   106 gagggccaagctgccaagctgggaccagggtcatgtcaggggcccggggcgcccagggcgc 165
          ||||||| | || ||| ||||| || || || || ||||| | ||||||| | |
Db   61 GAGGCCGGGCGGGGGCCCCGGGGCCCGGGGCGCCCGGGGG-CGCCACCGGGCCGGCTGGTGT 119

Qy   166 ggcgcactggcgcgggcgcgtgccgggagcctcctggtcgctctctgctctcgcgctccgc 225
          ||| ||| || |||||||| |||| | |||| || | | || | || ||||| |
Db   120 CCCCTCCTCGGCGCGGGCGGGAACTCCTCCTCAGGCCCTCGCGCTGCTCCGGGTCTGCCGGT 179

Qy   226 cgctccggctcgaggagggagagctgggtcggtcgtgtggaccacctagtgactatccg 285
          || |||| ||| | || ||||||||| |||| | || | | || | | | || |
Db   180 CTCTCCCGGACGCTGGACCCGCAACCTCCTCCAGGGCTCGGTGGTGGGACGACACTGTTACTC 239

Qy   286 gatagtggcacaatctaagaatatccgggaccaccaacacactgttgc'345
          || || | || ||| | || || |||| | || |||| ||| |
Db   240 GACAGTGGCAAAACCTTACARCCATCCACTATACCACTACCTTACTAGTACTGTGTCT 299

Qy   346 gaaagacaattacagtaccaaagggaaaagctgattctgaggttggagatgtggat 405
          || | | ||| | ||||| |||| ||| || || |||| || |||| ||| ||
Db   300 AAATGGGAGATTCCAGTAAAGACGGGAGAAGAATTCCCATCAAGTTCGGTGACTGTTGAC 359

Qy   406 atcgaatcccagaacctgtctctgactatcttctctca---------- 447
          |||||| |||||| | || | || | ||| | |
Db   360 ATTGAAGATTCTGATTATTGTCACCTTAATTACTTGAAGATTTTCAATGGAGATC 419

Qy   448 agctctcagatcaatatggtccatactgtgggagatgctgtccaagaaactcttg 507
          || ||||| || |||| || | || ||||| ||| | || || | || |||
Db   420 AGCAGAACCGGAAATAGGCACAGTACTGTCGGTGGGTCTCGGGAGATAGATAGAAGCAGGATGT 479

Qy   508 ttgaacacaagtgaagtaaccgtccgttgagagtggatcccacacattctgcggggggt 567
          || || | ||| || | || | || | || | || || | | | | |
Db   480 TCCAAAGCACGTGAAAAATCACACTGTCTTCATGAGTGCGGATCCAACGACTACAGTTGAC 539

Qy   568 ttttgctgcctatgccgagcagcgaccacctccagatttaatcacttgttggaacagagc 627
          ||||| || || || |||| || | || ||| | ||| | || | || || ||
Db   540 TTTTGGCTTCTTACTCAGTTATAGATAGTAAAAACAAGAGATTTAATCACTTGTTGGATCAGTA 599

Qy   688 gaggacacattctggggaatatggtggtagatcaacctcttttattgtcggaaagct 747
          || || |||| ||||| |||| || | || || | | | | || | || || |
Db   660 GCTGAAAATTCTGACAGATTCTCTCATGGATAATATATAGAAGGATATCTCACCCGTGTATGCCT 719

Qy   748 gccatccattgcagagaaatcatgctctgatgaactcaggtggccagagcagtgtcaagcgc 807
          || ||||| || |||| | || |||| ||||||||||| || || |||| || | ||
Db   720 GGAATCCATTGCAGGGATCAGTGTCTGATGAACTCAGGTGGGCTCCAGAGATGGGA 779

Qy   808 aaaggacagatcgaatcagagaatctgggacaatggtgttcttctggagggtaggttc 867
          || ||||| ||||| | |||| || || | ||||| ||| | |||| |||| || |
Db   780 AAAGGACACCCCATATACAGGGATCTTGGGCAACAATGTCACTTCACCTGAGTTC 839

Qy   868 ctgtcacgacaagcgagttctgttctaccttccaccatggttgcagagatcagattgaa 927
          || || || |||| ||| ||| ||| || || | || ||| | || | || ||
Db   840 TTATCACAGCAGTCGGTTCTGTTACATTTAAGACAAGTGGTTCCTATGGGCCTATGGGA 899

Qy   928 cctgacggg----------caatcagagctctctcctcatggcagtcgagtcggttcaatgag 975
          |||| |||| | || ||||| | | | |||| | | | | |||| ||
Db   900 TCAGGTGATCGCCGAYCCCCAGATAAACAGCATCATCYTTACGTCTAGGAGGTTCAATGAG 959

Qy   976 agtggagaccaagttcaactgtctcctggccaagaccacttcaggagcacaaggacccatca 1035
          || ||||| ||||| ||||| | ||| || ||| |||| || ||| || || || | |
Db   960 AGTGGAGACCAAGACCTGAGAAACCCGAGTAGCTGAGAAAACCGGGGCCGTCTCC 1019

Qy   1036 tgggcttcgggcgcacagtagcaacaaccaacaaacacagagtggctggagatggattg 1095
          ||||| | || | ||| | | ||||| | | | || | | || || | |||
Db   1020 TGGGCTTCGTTTTGCCACTGATGACGATC----------AGTCGCTGCGAAATTGACCTT 1067

Qy   1096 gggagaaaagaaaatacacagaattaggaccacagaactccacaacagtcgaactcaac 1155
          ||| | | | | ||| | | || || | | || | || ||||| | || || ||
Db   1068 AATAAGGACAGAGAGATTAACAGGCCATCGTAACCACTGTTCACCTGATTCTGATACACAAT 1127

Qy   1156 ttttatgttaaggagtttttgtgatgaacttcaaaaacaaataatctaagtggagagaccctat 1215
          |||| ||| |||| ||||| ||| || ||| ||| | || | || | |||| |||||
Db   1128 TACTATGTGTCTCGCCTACAGCAGTTCTGTTTGTTAACGGTGACGAGATGGGCAGAGGCTGTGTAC 1187

Qy   1216 aaaggaattgtgataataagagaaaaagtgttctcaggtaacctaacttccgggaccca 1275
          || |||| |||| | ||| | | |||| | | | | | || | | ||| |
Db   1188 AGAGAGCCTGGTGCCGCTGCAGGACAAGAGTAATATCAGATTATCAGAGGAT 1247

Qy   1276 gtgcacacaggtagagctctggccagatatgtgcgggtgttgtccccagaca 1335
          ||||| | ||| ||| | |||| | || || | ||| | ||| ||| |
Db   1248 GTTCGTGTTAACTTTTTGCCCACCAANTTATTGCACGTTTCATTAGAGTGAACCTTGTCCAG 1307

Qy   1336 tgtgcaccaaggtagctgttgaagtggcttgaaggtcattggttgccagattcac 1387
          |||||| |||| || | | |||| | |||||| | | || | || || | ||
Db   1308 TGGCAACACAGAAAATTGCCATGAAAGTGGATGAATTTTGCCGGGATGTGCAGTTCACTC 1359

IMPORTANT: This sequence is the entire insert of clone 94G16.
During sequence assembly data is inferred from overlapping clones.
Where differences are found these are annotated as variations
together
with a note of the overlapping clone name. Note that the variations
annotated may not be found in the sequence submission corresponding
to
the overlapping clone as we submit sequences with only a small
overlap
as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre chromosome 6
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6/
This sequence has been finished according to sequence map criteria
as follows. An attempt has been made to resolve all sequencing problems,
such
as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 94G16 is at 1 in this sequence. The true
right end of clone 94G16 is at 100808.
94G16 is from the library RPCI1 constructed at the Roswell Park
Cancer
Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.

```
FEATURES             Location/Qualifiers
     source          1..100808
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6q21"
                     /clone="RPI-94G16"
                     /clone_lib="RPCI-1"
     repeat_region   2..109
                     /note="AluSg repeat: matches 193..300 of consensus;
                     incomplete repeat"
     repeat_region   586..684
                     /note="MIR2 repeat: matches 32..134 of consensus"
     repeat_region   828..1128
                     /note="AluJb repeat: matches 1..300 of consensus"
     repeat_region   1511..1913
                     /note="MSTA repeat: matches 426..1 of consensus"
     repeat_region   2069..2366
                     /note="AluY repeat: matches 1..297 of consensus"
     repeat_region   2560..2589
                     /note="MIR repeat: matches 262..47 of consensus"
     repeat_region   4018..4220
                     /note="15 copies of 2 mer 100 % conserved"
     repeat_region   4380..4673
                     /note="MER42B repeat: matches 1204..1005 of consensus"
     repeat_region   5738..5790
                     /note="MER42c repeat: matches 983..690 of consensus"
     repeat_region   6052..6137
                     /note="MER5A repeat: matches 131..183 of consensus"
     repeat_region   7850..8066
                     /note="MIR repeat: matches 57..144 of consensus"
     repeat_region   9495..9676
                     /note="AluY repeat: matches 1..297 of consensus"
     repeat_region   11587..11690
                     /note="AluSx repeat: matches 119..300 of consensus;
                     incomplete repeat"
     repeat_region   12013..12188
                     /note="AluJb repeat: matches 1..302 of consensus"
     repeat_region   12310..12500
                     /note="L1MD1 repeat: matches 971..793 of consensus"
     repeat_region   13051..13336
                     /note="AluSp repeat: matches 108..299 of consensus;
                     incomplete repeat"
     repeat_region   13800..14096
                     /note="AluY repeat: matches 1..289 of consensus"
     repeat_region   16740..16877
                     /note="AluSx repeat: matches 302..3 of consensus"
     repeat_region   
                     /note="MIR2 repeat: matches 144..1 of consensus"

     repeat_region   17538..17591
                     /note="2 copies of 27 mer 100 % conserved"
     repeat_region   18151..18887
                     /note="FLAM_C repeat: matches 1..131 of consensus"
     repeat_region   18554..18722
                     /note="MIR repeat: matches 261..84 of consensus"
     repeat_region   20153..20294
                     /note="MER2c repeat: matches 1097..949 of consensus"
     repeat_region   21524..21613
                     /note="MIR repeat: matches 60..144 of consensus"
     repeat_region   21618..21915
                     /note="AluJb repeat: matches 1..301 of consensus"
     repeat_region   21916..22025
                     /note="MIR repeat: matches 141..262 of consensus"
     misc_feature    21983..21992
                     /note="Tn10 flanking repeat; Tn10 sequence was inserted at
                     position 21992..21993; in some sequenced clones which were
                     not included; in the final assembly"
     repeat_region   22793..22848
                     /note="28 copies of 2 mer 93 % conserved"
     repeat_region   23337..23527
                     /note="MIR repeat: matches 243..45 of consensus"
     repeat_region   26788..27074
                     /note="AluJo repeat: matches 1..299 of consensus"
     repeat_region   27487..27787
                     /note="AluSg repeat: matches 297..1 of consensus"
     repeat_region   28466..28524
                     /note="MIR2 repeat: matches 146..86 of consensus"
     repeat_region   28475..28537
                     /note="MIR repeat: matches 257..195 of consensus"
     repeat_region   28538..28843
                     /note="AluY repeat: matches 1..299 of consensus"
     repeat_region   29096..29432
                     /note="AluJo repeat: matches 354..1 of consensus"
     repeat_region   29433..30977
                     /note="THE1B-INTERNAL repeat: matches 1580..2 of
                     consensus"
     repeat_region   30980..31322
                     /note="THE1B repeat: matches 354..1 of consensus"
     repeat_region   32594..32810
                     /note="THE1A repeat: matches 354..1 of consensus"
     repeat_region   33663..34860
                     /note="MER30 repeat: matches 1..229 of consensus"
     repeat_region   34906..35188
                     /note="MER20 repeat: matches 218..2 of consensus"
     repeat_region   36667..36969
                     /note="AluJo repeat: matches 302..2 of consensus"
     repeat_region   37045..37345
                     /note="AluSq repeat: matches 299..1 of consensus"
     repeat_region   39246..39818
                     /note="MER33 repeat: matches 541..300 of consensus"
     repeat_region   41305..41604
                     /note="MLT1E repeat: matches 568..1 of consensus"
     repeat_region   42255..42299
                     /note="AluY repeat: matches 301..1 of consensus"
     repeat_region   42702..43037
                     /note="MADE1 repeat: matches 46..2 of consensus"
     repeat_region   44291..44516
                     /note="MER1B repeat: matches 337..1 of consensus"
     repeat_region   47198..47261
                     /note="MLT1F repeat: matches 133..40 of consensus"
     repeat_region   49349..49557
                     /note="2 copies of 32 mer 92 % conserved"
     repeat_region   49728..50116
                     /note="MIR repeat: matches 7..203 of consensus"
     repeat_region   49728..50104
                     /note="AluY repeat: matches 10..298 of consensus"
     unsure          50103..50103
     unsure          50110..50111
     unsure          50117..50111
     repeat_region   50140..50235
                     /note="FLAM_A repeat: matches 133..40 of consensus"
     repeat_region   51950..52234
                     /note="AluJo repeat: matches 1..297 of consensus"
     repeat_region   52255..52557
                     /note="AluJo repeat: matches 1..297 of consensus"
```

Left column

```
     repeat_region    53635..53786
                      /note="AluSq repeat: matches 1..303 of consensus"
     repeat_region    54479..54779
                      /note="L1PA13 repeat: matches 432..569 of consensus"
     repeat_region    55853..56121
                      /note="AluSx repeat: matches 3..302 of consensus"
     repeat_region    55844..55947
                      /note="AluSl repeat: matches 300..5 of consensus"
     unsure           57419
     repeat_region    57421
                      /note="MIR repeat: matches 216..116 of consensus"
     unsure           58835..58880
     repeat_region    58880
                      /note="AluSq repeat: matches 26..70 of consensus"
     repeat_region    59694..59896
                      /note="MER5A repeat: matches 14..250 of consensus"
     prim_transcript  <60654..>62080
                      /note="incomplete repeat"
     repeat_region    63844..64143
                      /note="AluY repeat: matches 301..2 of consensus"
     repeat_region    64314..64584
                      /note="AluSc repeat: matches 1..281 of consensus"
     repeat_region    66300..66538
                      /note="MIR repeat: matches 3..261 of consensus"
     repeat_region    67079..67373
                      /note="AluSc repeat: matches 1..295 of consensus"
     repeat_region    68397..68450
                      /note="MIR repeat: matches 50..102 of consensus"
     repeat_region    68690..69514
                      /note="L1PB3 repeat: matches 897..64 of consensus"
     repeat_region    69957..70250
                      /note="AluSq repeat: matches 1..303 of consensus"
     prim_transcript  <72733..74862
                      /note="match: multiple ESTs; match: T24499,AA069666,U82788
                      H10089,H21591; match: N23570,AA128011,W03600,N48006
                      N55225; match: R44125,R24821,N29055,N94008,N32368; match:
                      AA026810,N29184,W45579"
```

```
Query Match     12.4%;  Score 218;  DB 9;  Length 100808;
Best Local Similarity  100.0%;
Matches  218;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      901 ggttgacacagatccttgagtttgaaatccaatcaagaagttcttcctcaatgg 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    51085 ggttgcacgacgaatccttgagtttgaaactgacgggcaaaatcagagcttctttctctcgcatcg 51144

Qy      961 cagtcgttcaatgagagtggagaccaagttcactgtctccctgccaagtcgacttcag 1020
Db    51145 cagtcggtcaatgagagtggagaccaagttcactgtctccctggccaagtcgagacttcag 51204

Qy     1021 gaccaagtcgccatcatggccttcggccgacgagtacaacaacaaccacacgagagtcg 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    51205 gaccaagcgcccatcatcggcgtccgcgacagtagcaacaacaacacacgagagtcg 51264
Db    51265 ctggaatcgattgggggaggaaaaaaataaacagg 51302
```

Right column

```
            TITLE     The sequence of Homo sapiens clone
            JOURNAL   Unpublished
  REFERENCE 2  (bases 1 to 142616)
            AUTHORS   Waterston,R.H.
            TITLE     Direct Submission
            JOURNAL   Submitted (22-MAR-2000) Genome Sequencing Center, Washington
                      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                      MO 63108 USA
  COMMENT   On Oct 25, 2000 this sequence version replaced gi:9951643.

            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: Information
            -------------- Project Information --------------
            Center project name: H_NH0117013
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 139293 bases at least Q40
            Consensus quality: 140128 bases at least Q30
            Consensus quality: 140609 bases at least Q20
            Insert size: 141000; agarose-fp
            Insert size: 142145; sum-of-contigs
            Quality coverage: 5.86 in Q20 bases; agarose-fp
            Quality coverage: 5.84 in Q20 bases; sum-of-contigs
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 7 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the contigs are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
  FEATURES             Location/Qualifiers
       source          1..142616
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="6"
                       /clone="RP11-117O13"
       misc_feature    1..1680
                       /note="assembly_name:Contig4"
       misc_feature    1681
                       1680: contig of 1680 bp in length
                       1780: gap of unknown length
       misc_feature    1781..11282
                       11282: contig of 9502 bp in length
                       11382: gap of unknown length
       misc_feature    11383..20730
                       20730: contig of 9348 bp in length
                       20830: gap of unknown length
       misc_feature    20831..39951
                       39951: contig of 19121 bp in length
                       40051: gap of unknown length
       misc_feature    40052..67313
                       67313: contig of 27261 bp in length
                       67412: gap of unknown length
       misc_feature    67413..97063
                       97063: contig of 29651 bp in length
                       97163: gap of unknown length
       misc_feature    97164..142616
                       142616: contig of 45453 bp in length
                       /note="assembly_name:Contig5"
                       /note="assembly_name:Contig6"
                       /note="assembly_name:Contig7"
                       /clone="RP11-117O13"
                       /organism="Homo sapiens"
                       /chromosome="6"
                       vector_side:right
                       clone_end:T7
                       /note="assembly_name:Contig8"
                       /note="assembly_name:Contig9"
                       clone_end:SP6
                       vector_side:left
```

REFERENCE
  AUTHORS   Waterston,R.H.

LOCUS   AC026646   142616 bp   DNA   linear   HTG 17-NOV-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-117O13, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC026646
VERSION   AC026646.5 GI:10998991
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 142616)

```
RESULT 13

LOCUS       AC09428/C
DEFINITION  Rattus norvegicus clone CH230-101K6,
            ***, 58 unordered pieces.
ACCESSION   AC09428
VERSION     AC09428.2  GI:17973372
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 171300)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Ara,J.R., Banks,T.R., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,L.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,C., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,J., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,M., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.R., Jacobson,A., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L.J., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,W.,
            Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martinek,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,E., Pace,A., Payton,B.,
            Peery,D., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,M.,
            Thomas,S., Usmani,K., Vasquez,V., Vera,V., Villalon,D., Vinson,R.,
            Walir,R., Wang,M., Ward-Moore,S., Warren,J., Washington,C.,
            Watlington,J., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.P., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171300)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:16930835.
            -----------------  Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -----------------  Project Information
            Center project name: GHUB
            Center clone name: CH230-101K6
            -----------------  Summary Statistics
            Assembly program: Phrap; version 0.990329first call to
            findPhrapList
            Consensus quality: 138590 bases at least Q40
            Consensus quality: 147464 bases at least Q30
            Consensus quality: 155580 bases at least Q20
            Estimated insert size: 152696; sum-of-contigs estimation
            Estimated insert size: 152696; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/GenBank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 58 contigs. The true order of the pieces is
            * not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1       11285: contig of 11285 bp in length
            *     11286       11385: gap of unknown length
            *     11386       16925: contig of 5540 bp in length
            *     16926       17025: gap of unknown length
            *     17026       20478: contig of 3453 bp in length
            *     20478       20578: gap of unknown length
            *     20579       26656: contig of 6078 bp in length
            *     26657       26756: gap of unknown length
            *     26757       33936: contig of 7180 bp in length
            *     33936       34036: gap of unknown length
            *     34037       38968: contig of 4932 bp in length
            *     38968       39008: gap of unknown length
            *     39009       43965: contig of 4897 bp in length
            *     43965       44065: gap of unknown length
            *     44066       48815: contig of 4750 bp in length
            *     48816       48915: gap of unknown length
            *     48916       51807: contig of 2892 bp in length
            *     51807       51907: gap of unknown length
            *     51908       56443: contig of 4536 bp in length
            *     56444       56543: gap of unknown length
            *     56544       59759: contig of 3216 bp in length
            *     59760       59859: gap of unknown length
            *     59860       64034: contig of 4175 bp in length
            *     64035       64134: gap of unknown length
            *     64135       68670: contig of 4536 bp in length
            *     68670       68770: gap of unknown length
            *     68770       75472: contig of 6703 bp in length
            *     75473       75573: gap of unknown length
            *     75574       79856: contig of 4285 bp in length
```

```
ORIGIN

Qy  901  ggttgcagcagatcctgactttgacccgacccgaaatcagagcttcttcctccatgg  960
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141261  GGTTGCAGCAGATCCTTGAGTTTTGAACCTGACGGGCCAAATCAGAGCTTCTTCCTCCATGG 141202

Qy  961  cagtccgtcaatgagaatgagagaccaagtcaatcgtctcctgccaagcccgacttcag 1020
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141201  CAGTCCGTCAATGAGAGTGGAGACCAAGTTCACTGGTCTCCTGCCAAGCCCGACTTCAG 141142

Qy 1021  gaccaaggccatcatggggcttcgggcgacagtagcaaccaaccaccacacgagagtgg 1080
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141141  GACCAAGGCCATCATGGGGCTTCGGGCGACAGTAGCAACCAACCACCACACGAGAGTGG 141082

Qy 1081  ctggagatcgattggggggaggaaaaaaaataacagg 1118
                     ||||||||||||||||||||||||||||||||||||
Db 141081  CTGGAGATCGATTTGGGGGAGAAAAAAAAATAACAGG 141044

//
```

```
Query Match  97164.  142616
BASE COUNT   43950 a   26378 c   28311 g   43375 t    602 others
ORIGIN
misc_feature   97164.  142616
             /note="assembly_name:Contig10"

Best Local Similarity  12.4%;  Score 218;  DB 2;  Length 142616;
Matches 218;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy   1  ggccaaggccatcatggggcttcgggcgacagtagcaaccaaccaccacacgagagtgg  ...

LOCUS       AC09428   171300 bp   DNA   linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-101K6, *** SEQUENCING IN PROGRESS
```

```
 79859:
 79959:  83831:  contig of 3873 bp in length
 83832:  83931:  gap of unknown length
 83932:  86749:  contig of 2818 bp in length
 86750:          gap of unknown length
 86850:  89989:  contig of 3139 bp in length
 89988:          gap of unknown length
 90089:  93114:  contig of 3026 bp in length
 93115:  93214:  gap of unknown length
 93215:  95528:  contig of 2314 bp in length
 95529:  95628:  gap of unknown length
 99647:  99746:  contig of 4118 bp in length
 99747:  102546: gap of unknown length
102547: 104912:  contig of 2266 bp in length
104913: 105012:  gap of unknown length
105013: 108083:  contig of 3071 bp in length
108084: 108183:  gap of unknown length
108184: 111047:  contig of 2864 bp in length
111048: 111147:  gap of unknown length
111148: 112820:  contig of 1673 bp in length
112821: 112920:  gap of unknown length
112921: 115580:  contig of 2660 bp in length
115581: 115680:  gap of unknown length
115681: 117534:  contig of 1854 bp in length
117535: 117634:  gap of unknown length
117635: 119548:  contig of 1914 bp in length
119549: 119648:  gap of unknown length
119649: 121877:  contig of 2229 bp in length
121878: 121977:  gap of unknown length
121978: 124058:  contig of 2081 bp in length
124059: 124158:  gap of unknown length
124159: 126116:  contig of 1958 bp in length
126117: 126216:  gap of unknown length
126217: 128515:  contig of 2299 bp in length
128516: 128615:  gap of unknown length
128616: 130603:  contig of 1988 bp in length
130604: 130703:  gap of unknown length
130704: 132213:  contig of 1510 bp in length
132214: 132313:  gap of unknown length
132314: 134064:  contig of 1751 bp in length
134065: 134164:  gap of unknown length
134165: 136099:  contig of 1935 bp in length
136100: 136199:  gap of unknown length
136200: 137441:  contig of 1242 bp in length
137442: 137541:  gap of unknown length
137542: 139039:  contig of 1498 bp in length
139040: 139139:  gap of unknown length
139140: 140911:  contig of 1772 bp in length
140912: 141011:  gap of unknown length
141012: 142584:  contig of 1573 bp in length
142585: 142684:  gap of unknown length
142684: 144313:  contig of 1629 bp in length
144314: 144413:  gap of unknown length
144413: 144560:  contig of 2147 bp in length
144561: 144660:  gap of unknown length
146561: 146660:  gap of unknown length
146661: 148283:  contig of 1623 bp in length
148284: 148383:  gap of unknown length
148384: 150257:  contig of 1874 bp in length
150258: 150357:  gap of unknown length
150358: 151846:  contig of 1489 bp in length
151847: 151946:  gap of unknown length
151947: 153174:  contig of 1228 bp in length
153175: 153274:  gap of unknown length
153275: 155485:  contig of 2211 bp in length
155486: 155585:  gap of unknown length
155586: 156894:  contig of 1309 bp in length
156895: 156994:  gap of unknown length
156995: 158140:  contig of 1146 bp in length
158141: 158240:  gap of unknown length
158241: 159705:  contig of 1465 bp in length
159706: 159805:  gap of unknown length
```

RESULT 14

LOCUS   AC026646   142616 bp   DNA   linear   HTG 17-NOV-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-117O13, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC026646
VERSION   AC026646.5 GI:10998991
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 142616)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 142616)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   On Oct. 25, 2000 this sequence version replaced gi:9958143.

            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ------------- Project Information ---------
            Center project name: H_NH0117013
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid: 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 139293 bases at least Q40
            Consensus quality: 140128 bases at least Q30
            Consensus quality: 140609 bases at least Q20
            Insert size: 141000; agarose-fp
            Insert size: 142145; sum-of-contigs

Query Match   9.0%; Score 159; DB 2; Length 171300;
Best Local Similarity 79.1%;
Matches 189; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
                                          Pred. No. 2, 9e-20;

Qy   253 ggtgatgtcgtgtggaacacctagtgacttatcaggagtgagtggcacacatgacatcaagat 312
          ||||||| || | || |||||| |||||||| || |||||| |||| || ||| ||||| ||||
Db   9559 GGTGATGGCTGCGGGCACATAGTGACATCTCAGGAGACAGTGCACAATGACATCTAACAGAAT 9500

Qy   313 tatccggagacctacccaatcacacgtttgcgaaaaagacaattacagtacaaaggg 372
          ||||| || ||| ||||| ||| ||| || | |||||| ||| | ||| | ||||||
Db   9499 TATCCAGGGAGACTTACCCCATTACACCGTGTGTGAAAAGATCAATCACACTCCGAACAGGG 9440

Qy   373 aaaagacgattctgaggtgggagattggataccagacaacctgtgtcttctgac 432
          | |||||| ||| | ||| ||| ||| |||| |||||| |||| | |||| |||
Db   9439 AAGACACTTATTCCTCAGGCTGGGCGATTTGAACATTGACTCCAAGACCTGCCACTTCCGAC 9380

Qy   433 tatcttcctcttccaccagctcttcagatcaatatggtccatactgtggaagtatgactgt 491
          |||||| |||||| |||| | |||| |||||| || ||| | || || | ||| ||||| |
Db   9379 TATCTCCTCTTTAGCAGTGCGACAAGCATCCAGATCAGTATGGTAAGCAACGGGTATGGCGGGATCTGT 9321

RESULT 15

LOCUS   HUMCUB1
DEFINITION   Human mRNA for unknown product, partial cds.

LOCUS   HUMCUB1   1388 bp   mRNA   linear   PRI 07-FEB-1999

ACCESSION   D29810
VERSION   D29810.1   GI:704440
KEYWORDS   .
SOURCE   Homo sapiens cell line Li21 cDNA to mRNA, clone cub-1.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1388)
  AUTHORS   Shibata,T.
  TITLE   unpublished
  JOURNAL   Unpublished (1994)
REFERENCE   2 (bases 1 to 1388)
  AUTHORS   Shibata,T.
  TITLE   unpublished
  JOURNAL   Unpublished (1994)
COMMENT   On Mar 10, 1995 this sequence version replaced gi:474985.
             Submitted (08-Apr-1994) to DDBJ by:
             Tatsuhiro Shibata
             National Cancer Center Research Institute
             1-1 Tsukiji 5 chome
             Chuoku, Tokyo 104
             Japan
             Phone: 03-3542-2511 x4208
             Fax: 03-3248-2737.
FEATURES   Location/Qualifiers
     source   1..1388
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="Li21"
     CDS   <1..1097
                     /codon_start=3
                     /product="unknown"
                     /protein_id="BAA18909.1"
                     /db_xref="GI:704411"
                     /translation="DGCGIVLVGPESGCTLISLNXPQTYPNSTVCSWEIRVKMGERVRI
                     KFGDPDIEDSDSCHFNLKIYNGIIGVSRFEIGKKVCGLGLGIQINHIIESKRNEITLLFMS
                     GIHWSRGPLASYSVIDKGRSNLIFGHCIOFFWNNISSVSTAPAGCLLPARISGTIPH
                     GYRDSSPLCMAGVHAGVSNVSNTLGGQISVVISKIPYYESSLANNVTSVVGHILLQVFF
                     TFKTSGCYGTLGWESSGGDRGSSNNSITVLEWFDHGQENSWKPKKSQAEKTWTALGAF
                     ATDEYGWLQIDLNKKKTYGIITTGSTWVSIITWCLPTESCTVWMGRNGLCTESLVVE
                     QDKIFQGNKRLITTRWVRNNFLPFIARLLE"
BASE COUNT   432 a   289 c   302 g   364 t   1 others
ORIGIN

Query Match   7.5%;   Score 132.8; DB 9; Length 1388;
Best Local Similarity 50.1%; Pred. No. 1.8e-15;
Matches 581; Conservative 548; Mismatches 30; Indels 30; Gaps 9;

Qy   254 gtgatggctgtgaacccagtgacttatcaggatgagtggcaacaatctaagaatt 313
           ||||||||||||||||||||||||||||||||||||||||| |||| ||||||
Db   1 GTGATGGATGTGGACAACACTGTTACTAGGGCCCTGAGAGTGGAACCCTTACATCCATAAACT 60

Qy   314 atcccggaccaccacacacaacactgttgtcgcgaaaagacaaatctacaaagggga 373
           || ||||||| ||| |||||| ||||||| |||||| |||||||||
Db   61 ACCCAGAGACCTATCCCAACAGCACTGTGTTCGTAGGATAAGATCCGGTGTAAAGATGGGAG 120

Qy   374 aaagactgattctgaggttgggagagaattggatatcgaa------------tccc 415
           || |||| ||| |||||||||||||||||||||||||||| ||||
Db   121 AGAGAGTTCGCATCAAAATTGGTGACTTTGACATTGAAGACTTCGATTCTCGTCCACTTTA 180

Qy   416 agacctgtgttctgactatctcttcaccagctcctcagatcaatatggtccatact 475
           |||||||||| |||| | ||| ||| |||| |||| ||||||||||||
Db   181 ATTACTTGAGAATTTATAATAGTGGAAGTGCAGCAGCTCAGAAATNGGCAAATACTGTG 240

Qy   476 gtgggaagtatgatcgttcccaagaactctgtgaacacagtgaagtaaccgtcgct 535
           |||||||||||||||| ||||| ||||| ||||||| || |||||||
Db   241 GYCTGGGGTTGCAAATCAACCACATTCAATTGAATCACAAAGGCAATGAATCACATAACTGTG 300

Qy   536 ttgagagtggatccaccacattctcggcgggtgtttttgctgacctatcgagcagcgacc 595
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||

```
Db   301 TCATGAGTGGAATCCANTGTTTCTGGACGCGGGATTTTTGGCCTCATACTCTGTATAGATA 360
             ||||      | |||| || |||| || | |   || ||| |       |      ||
Qy   596 atc-cagatttaataacatgtttggaacgagctagcca-ttatttgaagcagaataacag 653

Db   361 AACAAGAGATCTAATTACTTGTTTGGACACTCGCATCACAANTTTTTTTGGAACCTGAGTTCAG 420
             || |||    || ||||| || | |||  |      |       |||||||  ||||
Qy   654 caaattctg-cccagctggttctgtagaacgtagcaggaagacattctgggaatatggtag 712

Db   421 TAAGTACTGCCCCACGCTGGTTGTCTCTCCTTTCTCTGAGATATCTGGAACANTTCCTC 480
         ||||||| |||||||||||||||| | | || |||||||  | |||||| |||||||
Qy   713 atggatatagaagatacctcctttattgtgcaaagctgccatccatgcaggaataattgctg 772

Db   481 ATGGATATATAGAAGANTTCCTCGCCATTGTGCATCGGCTTGGTGTTGCATCAGGAGTGTGTCAA 540
         |||||||| ||||||  ||||| |||||||||||  | |||||||| ||||||||| ||||||
Qy   773 atgaactaggctggccagaatcagtgtgtcttcagccgcaaagggatcagtcgatatgaaggga 832

Db   541 ACACGTTGGGCGGCCCAAAATCAGTGTTGTAATTAGTAAAGGTATTCCCTATTATNGAAAGTT 600
         ||||||| |||| |||||||||||||| ||||||||||||| ||||||||| ||  |||||||
Qy   833 tctctggccaatgtttgtgttctttcgagggatggttccctgtcaagacaagcggatttct---gt 889

Db   601 CTTTGGCTAACAACGTCACACATCTGTGGTCGGGACACTTGATACTACAAGTCTTTTTTTACAT 660
          ||  |||| ||  |||   || |||| | |||||   |||||  | ||||| |||||||| ||
Qy   890 ttacctccaatgtttgcagcagatccttgagtttgaacctgacggcaaatcaggagctt 949

Db   661 TTAAGACAACTGGANTGTTATNGGAACACTGGGNATCGGNATGGAGTCTGGTTGGTTGGTCGATCCTC 720
         || ||  | ||||  |||||| ||||||||||  ||||| ||||||||||| ||||||  |||||||
Qy   950 cttcctcaatggcagtcgtgtcaatgaagtggagacaagttcactggtctcctggcaag 1009

Db   721 CAAATANCAGCANTCACNCTGTGTCTGCGACTGACCACAGGGCAAGAGAACAGTTGGA 780
         |||| | ||||  | |   | | |||  ||||| ||||||||||||||||||||||||
Qy  1010 cccgacttcaggacacaaggcccatcatgggcttcgggcgacagtagcaacaccacaaac 1069

Db   781 AACCCAAAAAAAGCCAGGCTGAAAAAACCTGGACCCGCCTTGGGTGCNTTTGCCACTGATG 840
          |||||||||||  || ||||||||||| || || | || ||||||  |||||||| |||
Qy  1070 cacagaagtgtgtgagatcgatttcggggggagaaaaagagaaaataacagagaattaggacca 1129

Db   841 AATACCAGTGGTTACAAATAGANTTGANTANGGAANAAGANAAAANTAACAGGCANTTATNACCA 900
         || | | ||||| ||||| ||  |||| || ||||  || | |||  ||||||| | || |||
Qy  1130 caggatctacacagtcgaacttcaactcttatgttaagagtttgtgatgaacttcaaaa 1189

Db   901 CTGGANTCCNACCATGGTGCAGC-ACANTTACNTANTGTGTCNTGCGCTNACAGANTCCNTGTACNGTG 959
         |||| | || |||| |||| ||  ||  ||| | | ||||| ||  || |||| | || |||||| |||
Qy  1190 acaataanttcttaagtggaagaacctataaag-gaattgtgaataatgaagaaaaggtgtt 1248

Db   960 ATGATGGGCACGAAANTGGACTGNTGTNACAGNAGACCCTGGNTCCTGGNACGANTAAGANTANTTT 1019
         |||| ||| ||  ||  ||||| | || ||||  ||||| || |||||| | |   ||||  |  |||
Qy  1249 cagggtaactctaacttccgggac---ccagtgcaaaacaatcatccctccatcgtg 1305

Db  1020 CAAGGAAACANAAGANTANTCACCAGGATGGTGCCTANTAACTTTTTGCCCACCGANTANTT 1079
         ||||| ||| | || || || |||||| || |  ||| || |||||||||| || || |
Qy  1306 gccagatatgtgcggtgtctccccagacatggccaccagaggagatagccttgaagtggag 1365

Db  1080 G-CACGTTATTTAGAGTGANTCCTNACCCANTGGCNACGAGCAGCAGNAGAAANTTGCCATCNAAANTGGAG 1138
           |||||| || |||||| | |||  |||  ||| ||| |||||||| |||| |||||||| |||  ||||
Qy  1366 ctcatcggttgccagatta 1384

Db  1139 CTGCTCCGGANTGTCAGTTTA 1157
         |||| | ||  |||||||| |
```

Search completed: July 9, 2002, 01:13:21
Job time: 6038 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 23:26:13 ; Search time 1913.71 Seconds
                                               (without alignments)
                                               12419.927 Million cell updates/sec

Title: US-09-691-344A-3
Perfect score: 1761
Sequence: 1 atggaatccggtgcggggca............caagtgatatggcaggttaa 1761

Scoring table: IDENTITY_NUC
                       Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
                    1: em_estba:*
                    2: em_esthum:*
                    3: em_estin:*
                    4: em_estnu:*
                    5: em_estov:*
                    6: em_estpl:*
                    7: em_estro:*
                    8: em_htc:*
                    9: gb_est1:*
                   10: gb_est2:*
                   11: gb_htc:*
                   12: gb_gss:*
                   13: em_gss_hum:*
                   14: em_gss_inv:*
                   15: em_gss_pln:*
                   16: em_gss_vrt:*

Pred. No. is the number of results predicted to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	540.8	30.7	1924	11	AK016485	AK016485 Mus muscu
2	513.8	29.2	2835	11	AK014521	AK014521 Mus muscu
3	470.6	26.7	872	10	BG245910	BG245910 602358725
4	444.8	25.3	883	10	BG975603	BG975603 602845306
5	434.2	24.7	466	9	AW390343	AW390343 RC4-ST018
6	355.4	20.2	434	10	AW483656	AW483656 5586 MAR
7	335.4	19.0	406	10	BF740455	BF740455 RC3-HH000
8	273.2	15.5	275	10	BF144325	BF144325 601787144
9	252.2	14.3	257	10	BE068835	BE068835 MRO-BT037
10	249.2	14.2	260	10	BF331206	BF331206 MRO-BT037
11	228.2	13.0	474	9	AA050442	AA050442 mj12e10.r
12	228.4	13.8	883	10	BB860022	BB860022 BB860022
13	212.2	13.0	451	10	BP086926	BP086926 RC3-HT023
14	209.4	11.9	406	9	AA042037	AA042037 m04h12.r
15	203.2	11.5	406	9	AA042037	AA042037 m04h12.r
16	198	11.2	896	9	BF139411	BF139411 601785285
17	189.8	10.8	318	10	H32466	H32466 EST107591 R

RESULT 1
AK016485
LOCUS   AK016485   1924 bp   mRNA   linear   HTC 19-JAN-2002
DEFINITION   Mus musculus adult male testis cDNA, RIKEN full-length enriched
               library, clone:4931429A04:putative CUB domain containing protein,
               full insert sequence.
ACCESSION   AK016485
VERSION   AK016485.1 GI:12855243
KEYWORDS   HTC; CAP-trapper.
SOURCE   Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
               clone.lib:RIKEN full-length enriched mouse cDNA library
               clone:4931429A04.
ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99272953
PUBMED   10346636
REFERENCE   2 (sites)
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
               Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
               prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499974
PUBMED   11042159
REFERENCE   3 (sites)
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
               Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
               Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
               Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
               Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
               Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
               Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
               sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)

ALIGNMENTS

18	185.2	10.5	482	10	BP971665 QV0-ET015
19	179	10.2	219	9	BE068836 MRO-BT037
20	160.2	9.1	914	10	B1411107 602990241
21	155.6	8.8	295	9	BD585422 BD585422
22	150.6	8.6	643	9	BB179946 BB179946
23	148.8	8.4	608	10	BG257635 602377506
24	146.2	8.3	670	9	BB657747 BB657747
25	138.8	7.9	670	12	B57216 CIT-HSP-200
26	136.4	7.7	667	9	AL586285 AL586285
27	134	7.6	639	9	AW383890 QV3-HT036
28	129.8	7.4	639	9	AL048059 AL048059
29	128.2	7.3	587	9	AW383890 QV3-HT036
30	127.8	7.3	459	10	N46006 N46006
31	127.2	7.2	466	9	AI750960 yy35g05.r1
32	125.2	7.1	692	10	BF732861 BF732861
33	121	6.9	463	10	B1080802 B1080802
34	121	6.9	986	10	B1080802 B1080802
35	119.2	6.8	408	10	BF323384 BF323384
36	114	6.5	908	10	BF777080 BF777080
37	113.2	6.4	299	10	BF350722 BF350722
38	108.8	6.2	703	9	BF177683 BF177683
39	102.2	5.8	703	9	AW070902 AW070902
40	102	5.8	324	10	H84766 H84766 ys70e05
41	101	5.7	408	10	BF825585 BF825585 CM1-HN001
42	102	5.8	408	10	BF825585 BF825585 CM1-HN001
43	97.8	5.5	745	9	AA122907 AA122907 mg23a10.r
44	96	5.5	745	9	AL577036 AL577036
45	94.4	5.4	927	9	AL552613 AL552613

```
 MEDLINE   20530913
 PUBMED    11076861
REFERENCE   4  (sites)
 AUTHORS   ...FANTOM Consortium. The RIKEN Genome Exploration Research Group Phase II Team and the
 TITLE     Functional annotation of a full-length mouse cDNA collection
 JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1924)
 AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
           Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
           Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
           Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
           Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
           Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
           Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,N.,
           Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
           Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
           Schimil,L., Shibata,K., Shimada,H., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
           Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
           Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
           Hayashizaki,Y.
 TITLE     Direct Submission
 JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
 COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
           further details.
           cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues. First strand cDNA was primed with a primer
           [5'-GAGAGAGAGAGAGACAAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTVN 3'], cDNA was
           prepared by using trehalose thermo-activated reverse transcriptase
           and subsequently enriched for full-length by cap-trapper. cDNA went
           through one round of normalization to Rot = 10.0 and subtraction to
           Rot = 100.0. Second strand cDNA was prepared with the primer
           adapter of sequence [5'-
           GAGAGAGAGTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved
           with BamHI and XhoI. cDNA of size comprised longer than 7 kb was
           selected before cloning. Vector: a modified pBluescript KS(+) was
           bulk excision from lambda FLC I. Cloning sites; 5' end: SalI; 3'
           end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1..1924
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1894607"
                     /db_xref="taxon:10090"
                     /clone="4931429A04"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             133..1644
                     /note="data source:Pfam, source key:PF00431, evidence:ISS
                     putative
                     putative CUB domain containing protein"
                     /codon_start=1
                     /protein_id="BAB30265.1"
                     /db_xref="GI:12852244"
                     /translation="MGTGAGPSVIALLFAVCAPLRLQAEELGDCCGHIVTSQDSGTM
                     TSKNTPGTIPNYTVCEKIITYPKGKRLILRLGDLNIESKTCASDYLLFSSATDQYGPY
                     CGSWAVPEKLRLNSNEVTVLFKSSHISRGFLLTYASSDHPDLITCLERGSHYFEEK
                     YSKFCPACRDIAGDISGNTKDGYRSVNTQKAAHAGIITHGIENLLQSKGISH
                     YEGLLANGVLSRHGSLSEKRFLPTTPGMNTTTVAIPSVIFIALLTGHCIPA1CRKRKY
                     KGNPYVSADAQKTSCWKQIKYPPARIQSTEFTISYDNEKEMTQKLDLITISDMADYQQ
                     PLMIGTGTVARKGSTRPMDTDIEVANNTEASGHICPIRPGRHEYALPLTHSEPEY"
```

```
BASE COUNT      476 a    496 c    514 g    438 t
ORIGIN

Query Match             30.7%;  Score 540.8;  DB 11;  Length 1924;
Best Local Similarity   82.9%;  Pred. No. 2.7e-105;
Matches  617;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy          1 ATPIVERHLLRAHTFSTQSGTKVRVPGGPRPRTHKHSHSSSGGFPPATGATGVESSYQRPASPK
PVGGGYDKPAASSFLDSRDPASQSOMTSGGDDGYSAPRNGLAPLNQTAMTALL"

Qy        159 cagcacgggcagcgcacatggcgcgcggggctgccggttggctgctcgcggt 218
              |||||| ||| |||||||||||||||||||||||||||||||||||||||||
Db        123 CGGCGGGACCCATGGGGACGGGACCGGGCTGGTGGGCCCAGCTGTCCTGGCCGT 182

Qy        219 ctccgcccgcctcgggctcgaggtcggaggagactcggttgtggctgaggcacctactggac 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GTGTGCTCCGCCTCCGGGTTGCAGGCGGAGGAGGACTCGGGTGGCCACATAGTGAC 242

Qy        279 ttatcaggtatggcacaatgacatctaagaattatcccgggacctcaccaatcacac 338
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        243 CTCTCAGGACAGTGGCACACATGCACACACTAAGAATTATCCCAGGGACTTACCCCAATTACAC 302

Qy        339 tgtttgcgaaaagacaattacagctaccacaaagggggaaaagaaactcttgtgcgaacacaag 398
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        303 TCAGTATGCGTCCATAATGTCTGCGGAGTTGGCGTTCTTCCCAAAGAAACTCCCGGCCTCAAACTCAAA 398

Qy        399 tttggatatcgaatcccagcaccctgtgcttctcctgactatctctctcctctcaccagctcttcaga 458
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TTTGAACATTGAGTCCAAGACCTGCCCTTCTGACTATCTCCTCTTCTCCTCCGCAACAGA 422

Qy        459 tcaatacaggtccaatactgtgggaagtatatgactgttcccaaagaaaaaaacctctgttgtcgaacaag 518
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CTACGGCCAGCAGTGGCACATCCAGCATCAGGATGGCTACAGGAGCGTGAACACAGA 518

Qy        519 tgaagtaaccgtcccgtcttgagagtgggaagtcccaacttctgtgctgttctccaaagaaactcctggcagctgggtac 578
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 CGAAGTGACTGTCCTCTTCAAGAGTGGCGATCTCACATTCTCTGGCGCCGGGGTGGGCTAC 542

Qy        579 ctatgtcgcagcgacacatccagacaatcaatacaacatgtttgaaagagctagccattatt 638
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 CTACGCCAGCAGTGGACCATCCAGAGATTTAATTAAACCTGTTTGGAACGAGGGCAGGGCCATTATT 602

Qy        639 gaaacaggaatacacaaaatttctgtgccagctgtgtggtagaagacgtagcgagacagtctc 698
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 CGAGGAAAAATACACAAAATTTCTGCCCAGTGTCTCTAGAGACAGCATAGCAGAGATAATTTC 662

Qy        699 tgggaatatggtagatggatagaggatagagatatacactcttattgtcgaaagtgtcgcatccaatgc 758
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 TCGGAATACAAAAGATGGTTACAGAAGAAAAATACCTCTTATTATTGTGTCGCAAAGTCTCGAAAA 722

Qy        759 aggaataaattgctgatgaactaggtggtccagatacaggtgctccaggccgcaaaggggatcag 818
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 AGGGATCATCACGATGAACTAGGTGGCAGCAGCTAGGGTTCGGCCACAAAGGATAAG 782

Qy        819 tcgatatgaagggattctgccaatggtgtcttcttcaggaggtggtccctgtcaagaacaa 878
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 TCACTATGAGAGGATTCTGCGGCCAATGGTTGTCTTCTTCTTCCGGGATTGGTTCCGGAAAA 842

Qy        879 gcgattctgttacctccaatgg 902
              ||||||||||||||||||||||
Db        843 GCGATTCTTTTTACAACCCCAG 866
```

```
                    clone_lib:RIKEN full-length enriched mouse cDNA library
                    clone:4631413K11.
ORGANISM            Mus musculus
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE           1 (sites)
  AUTHORS           Carninci,P. and Hayashizaki,Y.
  TITLE             High-efficiency full-length cDNA cloning
  JOURNAL           Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE           10349636
  PUBMED            10349636
REFERENCE           2 (sites)
  AUTHORS           Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
                    Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE             Normalization and subtraction of cap-trapper-selected cDNAs to
                    prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL           Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE           20499374
  PUBMED            11042159
REFERENCE           3 (sites)
  AUTHORS           Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
                    Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
                    Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
                    Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
                    Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
                    Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
                    Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE             RIKEN integrated sequence analysis (RISA) system-384-format
                    sequencing pipeline with 384 multicapillary sequencer
  JOURNAL           Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE           20530913
  PUBMED            11076861
REFERENCE           4 (sites)
  AUTHORS           The RIKEN Genome Exploration Research Group Phase II Team and the
                    FANTOM Consortium.
  TITLE             Functional annotation of a full-length mouse cDNA collection
  JOURNAL           Nature 409, 685-690 (2001)
REFERENCE           5 (bases 1 to 2835)
  AUTHORS           Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,B., Arai,A.,
                    Arakawa,T., Baldarelli,R., Barbera,M., Brownstein,M., Bult,C.,
                    Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
                    Hara,A., Hayashi,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
                    Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
                    Kato,H., Kawai,J., Kojima,Y., Kondo,H., Kouda,M., Koya,S.,
                    Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
                    Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
                    Saito,H., Saito,R., Sakai,C., Sano,H., Sasaki,D.,
                    Schimi,Y., Shibata,Y., Shibata,Y., Shinagawa,A., Shiraki,T.,
                    Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
                    Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
                    Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
                    Hayashizaki,Y.
  TITLE             Direct Submission
  JOURNAL           Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
                    Physical and Chemical Research (RIKEN), Laboratory for Genome
                    Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
                    Genomic Sciences Center and Genome Science Laboratory in RIKEN
                    Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
                    Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp;
                    URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
                    Fax:81-45-503-9216)
COMMENT             Please visit our web site (http://genome.gsc.riken.go.jp/) for
                    further details.
                    cDNA library was prepared and sequenced in Mouse Genome
                    Encyclopedia Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in RIKEN
                    Division of Experimental Animal Research in RIKEN contributed to
                    prepare mouse tissues. First strand cDNA was primed with a primer
                    [5' GAACAGAGAAGGAAGACCCAAGAACGCTCTTTTTTTTTTTTTTTTTTVN 3']. cDNA was
                    prepared by using trehalose thermo-activated reverse transcriptase
                    and subsequently enriched for full-length by cap-trapper. cDNA went
                    through one round of normalization to Rot = 10.0 and subtraction to
                    Rot = 100.0. Second strand cDNA was prepared with the primer
                    adapter of sequence [5'
```

```
                                        GAGAGAGAGAATTCGGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
                                        with BamHI and XhoI. cDNA, of size longer than 7 kb was
                                        selected before cloning. Vector: a modified pBluescript K5(+) after
                                        bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3'
                                        end: BamHI. Host: DH10B.
FEATURES            Location/Qualifiers
     source         1..2835
                    /organism="Mus musculus"
                    /strain="C57BL/6J"
                    /db_xref="MGD:MGI:1894515"
                    /db_xref="taxon:10090"
                    /clone="4631413K11"
                    /tissue_type="skin"
                    /clone_lib="RIKEN full-length enriched mouse cDNA library"
                    /dev_stage="0 day neonate"
     CDS            74..1456
                    /note="data source:Pfam, source key:PF00431, evidence:ISS
                    putative
                    putative CUB domain containing protein"
                    /codon_start=1
                    /protein_id="BAB29409.1"
                    /db_xref="GI:12852431"
                    /translation="MSKNNPGTIYNNTVCEKIIITVPRGKRLLIRLGDLNIESKTCAS
                    DYLLFSSATDQYGPYCGSWAVPKELRLNSNEVTVLPKSGSSHSGRGFLLTYASSDHPD
                    LITLCLERGSHYFEERXSKFCPACRDRIAGDISGNTKDGYRDTSLLCARKIAGIITDE
                    LGHINLLQSKGISHVEGLLANGVLSRHGSLSERFLFTPGAMNITVAIPSVIFIAL
                    LLTGMGTFAICRKRKKGGNPYVSADAKTCKMQTIKYFARHQSTEFTISVDEKEMT
                    QKLDLITSDMADYQQPLMIGTGTVARKGSTFRPMDTDTEEVRVNTEASHCYDCPHRPG
                    RHEYALPLTHSEPEYATPIVERHLLRAHHPSTQSGYRVPGRPGPRTPHKHSHSSGGFPPAT
                    GAIQVESTQRPASPKPVGGGTDKPAASSFLDSRDPASQSQWTSGGDDGYSAPRNGLAP
                    LNQTAMTALL"
     polyA_signal   2816..2821
                    /note="putative"
     polyA_site     2835
                    /note="putative"
BASE COUNT          747 a           657 c           682 g           749 t
ORIGIN
```

```
Query Match     29.2%;    Score 513.8;  DB 11;  Length 2835;
Best Local Similarity   84.9%;  Pred. No. 1.8e-99;
Matches  575;  Conservative    0;  Mismatches 102;  Indels   0;  Gaps   0;

Qy         226 ccgctccggctgcaggcggagaggaggcgtgggtcggatggtcggtggaccactagtgactatcag 285
               |||||||||||||||||||| |||| |||| |||| |||| |||||||||||||||||||||||||
Db           2 GCCTCCGGTTGCAGGCGCGAGGAGCTGGGTGATGGTCGTGGCACATAGTGACCTCTCAG 61

Qy         286 gatagtggcacaatgacactctaagaattatcccggaggaccctacccaccaatcacactgtttgc 345
               |||||||||||||||||||| |||||| |||||||||| |||||||||||| |||||||||||||
Db          62 GACAGTGGCACAATGACANTCTAAGAATTATCCCAGAGGACTTACCCCAATTACACTGTGTCT 121

Qy         346 gaaaagacaattacagtaccacaaaggggaaaagagactgattcctgaggtttgggagagattggat 405
               |||||||||||||||||||||||||||||| ||||||||| |||||||||||||||||| |||||
Db         122 GAAAAGNTCATCACACTCCCAAAGGGGAGAGAGACTTATTCCTGAGGTTGCGGAGATTTGAAC 181

Qy         406 atcgaatcccagacgctgtgctctcgactatatctttcttcaccaagctcttcagaatgacaatat 465
               |||||||| |||||||| || |||||| ||||||| |||||||||||||||||||||||||||||
Db         182 ATTGAGTCCAAGACGCTGCCGCTTCTACTAYCTCCTCTTCTCTTCAGAGTGCAACAGATCAGTAT 241

Qy         466 ggtccatacgtggaagtctgactgttcccaagagaacttctgttgaacacaagtgaagta 525
               |||||||| |||||||||| |||| ||||||||||| ||||||||||||||||||||||
Db         242 GGTCCATAYTGTCGGAGTTCGGGCTGTTCCCAAGAGAACTCCGGCTCAACTCAAACGAAGTG 301

Qy         526 accgtccgcttgaaggtggatcccaccattcctcgccgcgggtttcttttgctggaccatgcg 585
               |||||| ||||| |||||| ||||||| |||||| ||||||||||||||| |||||||||| ||
Db         302 ACTGTCCTCTTCAAGAGTGGATCCACATTTCTCGGCCGCGGGGGGGGGCTTTCGTCTACGCC 361

Qy         586 agcagcgacacatccagattcaataacatgtttgaacgaggctagccattatttgaagaca 645
               |||||| |||||||||||||||||||||||| |||||||||||||||||||||||| |||||
Db         362 AGCAGTGACCATCCACAATTATAAATAACCTGTTTGGAACGAGGCCAGCATTAATTTCGAGGAA 421

Qy         646 gaatacagacaaattctgccaagctggttagagacgtagcagaagacacattctgggaat 705
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         705
```

```
Db   422  AAATACAGCAAATTCTGCCCAGCTGGCTGTGTAGAGACATACAGCAGAGAGATATTTCTCGCAAT  481
           ||||| |||  |||||||||||||||||||||||||||| |||||||||||||||||||||||||
Qy   706  atggtagatagatagagatacccctttatatttgttgcaaagctgccatccatgtcaggaata   765
Db   482  ACAAAAGATAGGTTACAGAGANTAACCTCTTTATTGTCAAAGCTGCCATCCACGGCAGGGAAYC  541
           |||||| ||| ||  ||||||||| |||||||||||||||||||||||||||||||||||||||
Qy   766  atcgtctgatgaactaggtggccagatcagtgtgcttcagcgcgacaaagggatcagtcgacat  825
Db   542  ATACACAGTACAGAACTAGGTGGCCACATCACAACTTGCTTCAGAGACAAAGGGATAAGTCACTAT  601
           ||  || |||||||||||||||||||| ||||||| ||||||||| ||||||||||| |||||
Qy   826  gaagggatcctcggccaatggtgtcttcttcgagagggatggtccctctcagacaagccgatt   885
Db   602  GAAGGACTCCTGGCCAATGGGCGTGTCCTCTCCCGGCAYGGTTCTTTGTCGGAAAAAGCCATTT   661
           ||||| ||||||||||||| | ||||| |||||||||  ||||||||||||||||||||| |||
Qy   886  ctgttacctcccaatgg   902
Db   662  CTTTTTTACAACCCCAGG   678
```

RESULT 3

```
LOCUS       BG245910               872 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602358725F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4486950 5',
            mRNA sequence.
ACCESSION   BG245910
VERSION     BG245910.1  GI:1275725
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 872)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)?
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10331 row: b column: 07
            High quality sequence stop:786.
FEATURES             Location/Qualifiers
     source          1..872
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4486950"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT    213 a    202 c    249 g    208 t
ORIGIN
```

```
Query Match   596;   Score 470.6;  DB 10;  Length 872;
Best Local Similarity   79.4%;   Pred. No. 2.4e-90;
Matches  596;  Conservative  0;  Mismatches 149;  Indels 6;  Gaps 3;

Qy   159  cggcggcggcggcactggcgcggggggcggggccctccggtgcttggctcgcggt   218
           ||||||  ||| ||||||||||||| | ||||||||| ||||| | || |||||
Db    49  CGGCGGGACCATGGGGACCGGGGCTTGGTGGGGCGGCAGTGTCCTGGGCGCTGTTCCCGCCT  108
Qy   219  ctccgcgcccgtccgcgcgcggcaggcgggcgggaggcgaggaggtcgatggctgtggaacctaagtgac  278
```

RESULT 4

```
LOCUS       BG975603               883 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  6028430F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4981009 5',
            mRNA sequence.
ACCESSION   BG975603
VERSION     BG975603.1  GI:14363240
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
Db   109  GTGTGTCTCCGCTCCGTTCAGGCGGTTGCAGGCGGAGGAGCTGGGTGATGGCGTGTGGGCACAGTAGTGAC  168
           ||||||| || | ||||| |||| ||||| |||| |||||| || || ||||||||||| || | |||||
Qy   279  ttatcaggatgatgtggcacacatgacatctctaagaaatattccccgagggacctaccccaatacac   338
Db   169  CTCTCAGGACAGTGGCACAACATCTTTATTGTCAAAGCTGCCATCCACGGCAGGGAATCAAAATTTACAC  228
           ||| |||||| || |||| ||||||||||| ||||||| |||||||||||||||||||||||||||||||
Qy   339  tgttgcgaaaagacaattaccgatacctaccacaaaggggaaaagaactctgaggtctggagga       398
Db   229  TGTGTGTGAAAAAGATCACAGATCCTCACAGTGGGGAAGACATCTGAGGTCTGGAGGAGAACTCTGAGGA  288
           || ||| |||| ||||| || ||| |  |||||||||  |||||||||||||||||||||||||||||||
Qy   399  tttggatcagaatcccacaggacctgttctcctcggacatcttctcaccacagtctcctcaa           458
Db   289  TTTGGACATAGATACCCACTCTGCTCCTCCTTTATGTCGCAAAGCTGCCATCCACGGCAGGGAATCAAAA  348
           ||||| ||| ||| |||||  ||||| |||| |||| |||||||||||||||||||||||||||||||||
Qy   459  ccaatatgttccaactttgtagcagagaagactcttgttgtggaacttgttgtgaacacaag           518
Db   349  CCAGATAYGTTGGTTGGGGACTGTTCCCAAAGAACTCCGGCTGAACTCAAA                     408
           ||||| | ||| | || | |  ||| |||| ||||||||  || |||| |||
Qy   519  tgagtaaccgtcccgctttgagagtggatcccacacttctgtgccggggtttttttgctgac         578
Db   409  CGAACTGACTGTCTCTCTTCAAGACAGGATTCACAGTTTCTCGGAGGACGTTCCGCATTCTGCGAC      468
           || |||||||||||| |||| || ||||||||||||||||||| ||||| || | | | ||  | |
Qy   579  ctatgcgagcagcgaccatccagaattcaataacatgtttggacagagacgagctagccagttatatt   638
Db   469  CTACCCCAGCAGTGACATCCAGCATTAATAACCTGTTTGGACAGAGACGAGCTAGCCAGTTATATT      528
           |||| || ||||  |||||||||||  |||||| ||||||||||||||||||||||||||||||||
Qy   639  gaagacagaatacagcaaatctgccccagcctgttgtagagatagcgaggagacattc             698
Db   529  CCAGGAAAAATACAGCAAATTTCGTCCCCAGCTGGCTGTGTAGAGACATAACAGGAGAYATTTC       588
           | |||||||||||||||||| |  |||||||| ||||||||||||||||   || ||| |||||
Qy   699  tgggaatatgattgatggtataagatacctcttat-tgtcgcaaagctgccatccatg            757
Db   589  TGGGAATTACAAAAGANTGGTTACAGAGANTACCTCTTTATTGTCGCAAAGCTGCCATCCACGG       648
           ||||||| || ||||  |||||||||||  ||||| | ||||||||||||||||||||||| ||
Qy   758  caggaataattgctgaataactaggtt-ggccagatcagtgtgcttcagcgcacaaaggg--      813
Db   649  ACGGGATCATCACAGATGAACTAGGTGGCCACATCACAACTTGCTTCAACTGCTTCAGCAGAAGGGGAT   708
           |||| || |||||||| ||||||||| |||| |||  |||| ||||||  || || ||| | |||  |
Qy   814  atcagtcgatatgaaggattctggccatgtgtgtcttcttcaagggggatggtcctcgtc           872
Db   709  AAGTCACCGTNTTGAAGGACTCCTGGCCAATGGGTGTCTTCTCCCGGCCATGGTTCCCTTGTC        768
           || ||| |   ||||||||  ||||||||| ||| |||||| |  || ||||||| || |||
Qy   873  agacaagccgattctgttaccctccaatggt                                        903
Db   769  GGAAAACGCGATTCTTTTTACAACCCCAGGAT                                       799
```

```
                http://image.llnl.gov
                Plate: LLAM1098l row: 1 column: 02
                High quality sequence start: 14
                High quality sequence stop: 750.
FEATURES        Location/Qualifiers
    source      1..883
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone_lib="NIH_CGAP_Mam4"
                /clone="IMAGE:4981009"
                /tissue_type="tumor, gross tissue"
                /lab_host="DH10B"
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Site_1: NotI.
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                Average insert 2.5 kb construced by Life
                Technologies, catalog # 12018-016. Investigators providing
                samples: Luthar Hennighausen/Priscilla Furth, NIH
                Reference for transgenic model: Li et al., Cell Growth and
                Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
                library."

BASE COUNT      230 a     197 c     255 g     201 t
ORIGIN

Query Match          25.3%;    Score 444.8;   DB 10;    Length 883;
Best Local Similarity 78.3%;    Pred. No. 7.9e-85;
Matches   597; Conservative    0; Mismatches 157; Indels    8;   Gaps    5;

Qy    159 cggaggcggcgcgcgcaactcccggccgcggggccctcctggcttcgtcggggt 218
          ||||||||    |  |||| ||||  ||||  || || |||||||| |||||||||
Db     73 CGGCGGGACCATGGGGACGCGGGGGCTGTGTGGGCGGCGACTGTCTGTCTGCCGT 132

Qy    219 ctccgcccgcctccggcctgcgcgggaggggcctgggtgatggctggacacccagtgac 278
          |||||| ||||| ||||| ||||| |||  ||||||||||||| |||||||||||||||
Db    133 CTGTGTCTCGGCTCCGGTTGCGGGAGGCAGCTGGGTGGTGATGGCTGTGGGCACATGGAC 192

Qy    279 ttatcagggataggtggcacatccaagaattatccgcggcgacctcacccacaacac 338
          ||| ||| ||||  |||||| |||||||||||| ||  |||||| |||  ||||||||
Db    193 CTCTCAGGGACAGTGGCACATCCTAAGAATNTATCAGGGACTTACCCCAATTACAC 252

Qy    339 tgtttgcgaaagacaattacagtaccaaaggggaaagactgagttctgaggaga 398
          || ||||  ||||  || || || |||||| |||||||  | |||||||||||||
Db    253 TGTGTGTGAAAAGANTCATNCACAGNGTCCCAAAAGAGACNGTTNTNTCGAGGAGA 312

Qy    399 -tttggatatcggaatccccagacactggttctgtagagacgtagcacaggagacatt 457
           ||||||| | |||||||| | ||| |||||| |||||| |||| ||| ||||| |
Db    313 TTTTGAACATTGAGTCAAGAGAACCTCGCCCGGCCTTCGACTANTCCTCTTCAGCACATGCAACAC 372

Qy    458 ataaatagtgtccatactggaagtatgactgttccaaagaactctgtgaacaca 517
          |||||  | |||| ||| |||| | | ||| |||| |||||| |||||||||||
Db    373 ATCAGTANTGTCAANATTGTGGGAGCTGGGCTGTTCCAATCAANAACTCCGGCGAACTCAA 432

Qy    518 gtgaagtaaccgtcgcttgaggtggatccacacattctgaggttgctga 577
          || | ||| ||| ||||||||  | ||||| |||| ||||||||| |||
Db    433 ACGAAGTGGACTGTCTCTTCTTCAAGCAGTGGGATCTCACANTTTCTGCCNGGGGGGCTTTCGTCTGA 492

Qy    578 cctatggcagcgcggccagagatta-ataacatgttggaacgagctagccattat 636
          |||| ||| |||| ||||| |||||| ||||||||| | || ||||  ||||||
Db    493 CCTACGCCAGCAGTCACCATCCAGTTACTATNAACCTGTNTTGNAACGNGGNGCCATNAT 552

Qy    637 ttgaagacgaataacgcaaattcgcccagcactggttctggtagcagggagacatt 696
          |||  ||  |||| || |||||||| |||||| ||||||| |||||| || ||
Db    553 TTCGAGGAAGAAAATACAGCAAATTCTGCCCAGCACTGGTTCNTGGTAGCACAGAGATATG 612

Qy    697 tctgggaaatatggtaggatgatta---gagatacctcctttattgtcaaagctgccatc 753
          ||||||||  ||| ||  || ||| |    |||||||||| ||| ||| |||||||||||
Db    613 TCTGGGAAATACAAAAGCATTGGTTACAGAGAAATACCTCTTTATTGTCAAAGCTGCCAT 672

Qy    754 catcaggaaataattgctgatggaactaagttactgctgctcagcgcaaagg 813
          |||| |||  |||| ||||| ||  ||||| ||||||| ||||| |||||||
Db    673 CACGCAGGAGTCATCAGANTGANTGAGGAACTAAGTAGCTGGACAACATCAAGANTGACATC 732

Qy    814 atcagtc-gatataagggaattctggccaa--tggtgttcttcgaggatggttccctg 870
          ||||||  || ||||||||||||||||||   |||||||||| ||||||||||||||
Db    733 ATCAGTCAGTTATNGAAGGAGCACTAGCTGGGGCAAAATTCAGNGTTCTCCCGCANGGTTCCNATG 792
```

```
Db    733 ATCAGTCAGTTATGAAGGAGCACTAGCTGGGGCAAAATTCAGGGTTCTCCCGCGANGGTTCCNATG 792
          |||||| ||||||||||||||  |  || |||||| ||  || ||| | ||| |    |  ||
Qy    871 tcagacaagcgattctgttacctccaatggtgtcagcagaga 912

Db    793 TGGGAAANGTGCCGATAACTNATGAAAAACCCAGGAAGTNAATAATGA 834

RESULT 5
LOCUS       AW390343                      466 bp    mRNA    linear
DEFINITION  RC4-ST0185-131099-011-c01 ST0185 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW390343
VERSION     AW390343.1  GI:6895002
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
  TITLE     The FAPESP/LICR Human Cancer Genome Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti-RC4&t2-RC4-ST0185-
            131099-011-c01&t3=1999-10-13&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 26
            High quality sequence stop: 204.
FEATURES    Location/Qualifiers
    source  1..466
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="ST0185"
            /clone="ST0185"
            /dev_stage="Adult"
            /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
            Site_2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716. Ludwig Institute for Cancer Research).
            profiles into the pUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."

BASE COUNT      120 a     116 c      96 g     134 t
ORIGIN

Query Match          24.7%;    Score 434.2;   DB 9;    Length 466;
Best Local Similarity 97.0%;    Pred. No. 1.3e-82;
Matches   453; Conservative    0; Mismatches  13; Indels    1;   Gaps    1;

Qy    809 aagggatcaatcgatataagggaggattctggccaatggtgtctcttcgaggatggttccc 868
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    466 AAGGGATCAATCGATATAAGGGAGGATTCTGGCCAATGGTGTCTCTTCGAGGATGGTTCCC 407

Qy    869 tgtcagacaaagcggattctgttacctccaatggtgtcagcaggtctctgagttcagaac 928
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 TGTCAGACAAAGCGGATTCTGTTACCTCCAATGGTGTCAGCAGGTCTCTGAGTTCAGAAC 347

Qy    929 ctgacgggcaaatcagaacttcctcctcagtcgtcaaatgagagtggagaccaag 988
          |||||||||||||| |||||||||||||||||||||||| |||||||||||||||
Db    346 CTGACGGGCAAATCAGAACTTCCTCCTCAGTCGTCAATGGAGAGTGGAGACCAAG 287

Qy    989 ttcactggttctctgccaagccgacttcaggaggacccatccatgggcttcgggcg 1048
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    287 TTCACTGGTTCTCTGCCAAGCCGACTTCAGGAGGACCCATCCATGGGCTTCGGGCG 1048
```

RESULT 6

```
LOCUS       AW483656                434 bp    mRNA    linear   EST 25-APR-2001
DEFINITION  55586 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW483656
VERSION     AW483656.1  GI:7053762
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
            G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCACTACCACG
            Plate: 28  row: E  column: 3
            Seq primer: ATTTAGGTGACACTATAG
FEATURES             Location/Qualifiers
     source          1..434
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCWV SPORT6; Site 1: XbaI; Site 2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      102 a     93 c     128 g     111 t
ORIGIN
```

Query Match   20.28; Score 355.4; DB 9; Length 434;
Best Local Similarity   90.38; Pred No. 8.9e-66;
Matches 360; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy     286 TTCACTGGTCTCCCTGCCCAAGCCGCGACTTCAGGACCCACATCATGGGCTTCGGGCG 227
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    1049 acagtagcaccaacaccaaccacagagagtgtgctggagtacgatcgttcgggggagaaaga 1108

Qy    1049 acagtagcaccaacaccaaccacagagagtgtgctggagtacgatcgttcgggggagaaaga 1108

Qy     226 ACAGTAGCAACTACCACTAACCCACCGACGAGTGGCTCGGACGAATCGATATTVGGGGAGAAAAAGT 167
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1108 ACAGTAGCAACTACCACTAACCCACCGACGAGTGGCTCGGACGAATCGATATTVGGGGAGAAAAAGT 73

Qy    1109 aaataacagagaattaggagaccaaggatcctacacacgtctcaacactttatgtctaaga 1168

Qy    1169 gttctgatcgaactccaaaaaacaaataaattctcatagtcggaagagacataaaggaaattgtga 1228

Qy     166 AAATAACAGAGAATTAGGAGACCAAGGATCCTACACACGTATTYACACGTCGAACTTCAACTTTTATGTCTAAGA 107

Qy     106 GTTTTCGTGATCGAACCTGAAAAAACAAATAAATTCTCATAGTCGGAAGACACCTATAAAAGGAATTGTGA 47

Qy    1229 ataatgagagaaaagtgttctcaggtaactctaacttccgggaccca 1275

Qy      46 ATAATGAGAGAAAAAGTGTTTCAGGGT-ACTCTAACTTCCGGGACCA 1
```

RESULT 7

```
LOCUS       BF740455                406 bp    mRNA    linear   EST 10-JAN-2001
DEFINITION  RC2-HB0006-051200-011-b07 HB0006 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF740455
VERSION     BF740455.1  GI:12067131
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/getehtml2.pl?ti-RC2&t2-RC2:HB0006-
            051200-011-b07&t3-2000-12-05&t4-1)
            High quality sequence start: 47
            High quality sequence stop: 406.
            Seq primer: puc-18 forward
FEATURES             Location/Qualifiers
     source          1..406
                     /organism="Homo sapiens"
```

```
Qy     613 tgtttggaacgagctagccattatttgaaggacgagacagagacacagcagaaattctgcgcggt 672
Db      14 TGTTTGGATCGATGCTAACCATTATTYAGAAGACAGAATACACAGGGAAGTTCTGCCCAGCTGGT 73

Qy      14 TGTTTGGATCGATGCTAACCATTATTYAGAAGACAGAATACACAGGGAAGTTCTGCCCAGCTGGT 73

Qy     673 tgtagagacgtagcaggagagacatctggtatagatggtagagatggagatgatacctct 732

Qy      74 TGTAGAGACGTACCAGGAGATATTYCTGGGAATAGTGTGGTGTGAATTCGGTAGATGGATTCTCT 133

Qy     733 ttatttgtcaagctgccatcacatcgcaggaaataatcgtcgatcgaactcagatc 792

Qy     134 TTATTTGTTGCAAGACGCCCATCCATCGCAGGACGTCAGGATGAACGTGGGTCAGATC 193

Qy     793 agtcatgttcatccacgcgcaaaagggatccagtcagtatcgataagtcgaagggaattctcgtgttctt 852

Qy     194 ACGGTGTTCTCGCCGCAAAAGGCCTAAGTCGCAATAGTAGGAGAAATCCTGGCCAATGGTGTTCTC 253

Qy     853 tcgaggcgatggttcccctgtcagacagacgatttctgttcaccctccaatggttcagccaga 912

Qy     254 TCAACGGCATGCGTTCCCTGTCTCCCGTCCTGGCCACAGACAAGGCGATTYCTGTTTACCTCCAATGGCTTGCACTGCGA 313

Qy     913 tccttgaagtttgaacctgacgggcgaaaagggttctctctcctcactgggcagtcggtcaat 972

Qy     314 TCCTTGACTTTGACGACACTTGACAGGAGAGCTTCTTCTCCTCTGGCGAGGATCAAT 373

Qy     973 gagagtggagaccaagttcactggttcactcctcggccaagcccgacttcaggaccaggccca 1032

Qy     374 GAGAGCTGGAGAATCAGGTCCCGCTGGTTCCTCTCGGCCAAGCCCCGACTTCAGGACCACCAAGGCCG 433

Qy    1033 t 1033

Qy     434 T 434
```

RESULT 8

LOCUS   BF144325   975 bp   mRNA   linear   EST 24-Oct-2000
DEFINITION   601787I44F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4014778 5',
             mRNA sequence.
ACCESSION   BF144325
VERSION   BF144325.1 GI:10983365
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 975)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov.
             Plate: LLAM9259 row: h column: 11

BASE COUNT   135 a   80 c   101 g   90 t
ORIGIN

Query Match   19.0%; Score 335.4; DB 10; Length 406;
Best Local Similarity   94.9%; Pred. No. 1.7e-61;
Matches 369; Conservative 0; Mismatches 16; Indels 4; Gaps 2;

Qy 1372 ggttggccagattacacaaagtaattgattccattggttgtggccgcagacaagtcaaagcacc 1431
         ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGAGGCATAGTTATCAAGGGTAATGATTCATACAGGTGTGG-GCAAGACAAGTCAAAGCACC 59

Qy 1432 agtgttcaactaagaaagaagatgagacaatacacaagggtcctgccatccccctcggagaaaaca 1491
         ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   60 AGTGTCTCAACTAAGAAAGAAGAGATGAGACAATACACAAGGGTCCCATCCCCCTCGGAGAAAACA 119

Qy 1492 tccacaggaataacattacacaacgtgtgctatccataaccaagggtcgctcctcctgtgtcctggtg 1551
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   120 TCCACAGGAATAAACATTACAACGTGTGGCTATTCCAATGGTGTCTCCTCTTGTGTAGTCCTCGTG 179

Qy 1552 tttgctgggatagggtacatcttgcagcctttagaagagaagagaaaggagggccgtat 1611
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   180 TTTGCTGGGAATCGGGATACCTTTGCACGCCTTTTAGA---AAGAAGAAGAAGGAGGGCCGTAT 236

Qy 1612 ggatcagccggaggctcagaaaacagactgttggaagcagagattaaataatatcccttgccaga 1671
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 GGATCAGCAGAGGGCTCAGAAAACAGACTGTTGGAAGCAGAGATTAAATATCCCTTGCCACA 296

Qy 1672 catcagtcagctcgagcttaccacccatcagtcatgtaatgagagaagagaagtgaaccacaaagtta 1731
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 CATCAGTCAGCTCGAGTTTACCATCACAGTCAGTAATGAGAAGAGAAGGAGATGACACAAAAGTTA 356

Qy 1732 gatctcatcacaaatggatatgccaggtta 1760
         ||||||||||||||| |||||||||||
Db   357 GATCTCATCACAACAAGTGATAATGGCCAGATTA 385

FEATURES   Location/Qualifiers
     source   1..975
                     /organism="Mus musculus"
                     /strain="C2ECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4014778"
                     /clone_lib="NCI_CGAP_Lu30"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; transgenic model WNT-1, expression driven by
                     MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
                     dT. Library constructed by Life Technologies,
                     Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT   248 a   240 c   299 g   188 t
ORIGIN

Query Match   15.5%; Score 273.2; DB 10; Length 975;
Best Local Similarity   82.4%; Pred. No. 4.2e-48;
Matches 361; Conservative 0; Mismatches 73; Indels 4; Gaps 4;

Qy   466 ggtccatactgtggaagtatgacatgtccccaaagaacaactctgttgtgaacacaagtgaagta 525
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   200 GGTCCATATAGTGGGAGTGGTTGGGCTGTTGTCCCAAAGAACAACTCCGTGGCTGAACTCAAACGAACGTG 259

Qy   526 accgtccgctttgagagtggggatcccacattcctggccgaggttttttttttgctgacctatgcg 585
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   260 ACTGTCGCTTCTTCAAGAGTGGAGTCAGATCTCACATTTCTGGCCGCTGGAGGTTTTCGTGCTGACTACCC 319

Qy   586 agcagcgaccatcccagattaatacaacatgtttggaacgagcgaagctagccattattgaagaca 645
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   320 AGCAGTGAACCATCCCAGATTTAATAACCCTGTTTGGAACGAGCGAGGCTAGCCACATTATTGCAGGAA 379

Qy   646 gaatacagcaaatcctcgccaagctcgttgtagagagcgtagagagagaagggacacattctgtgggaat 705
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   380 AAATACAGCAAATTTCTGCCCAGCTGCGTTGAGAGCAGTAGCAGAGATAATTCTGCGGAAT 439

Qy   706 atcgtagatcggataagatacctctttattgtcaaagcgcgcatccaatccgaggaata 765
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   440 ACAAAAGATGG-TACAGAGANTACCTCTTTATTGTGCAAAGCTG-CATCCACGCGACGGATC 497

Qy   766 attgctgatacaactaggtgtgcca--gatcagtgtgctccagacaaaggggatcacagtcgata 824
         || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   498 ATCACAGATGAAGAACTAGGTGGCGCCACCATACAACTTCGTTCAGACACACGGGATNAAGCTACTA 557

Qy   825 tgaaaggattctggccaatggtgttctttcaggatggtggttccctgccagacaagacgagatt 884
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   558 TGAAGGACGTCCTGG-CAATGGCGTGCTCTCCCGCGNCATGGTCTTGTCGGAAAACGGATT 616

Qy   885 tctgcttcacctccaatgg 902
         ||| |||||||||
Db   617 TCTTTTACACCCCAGG 634

RESULT 9

LOCUS   BE068835   257 bp   mRNA   linear   EST 09-JUN-2000
DEFINITION   MR0-BT0374-220300-001-f07 BT0374 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE068835
VERSION   BE068835.1 GI:8413485
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 257)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
             Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

---

ORIGIN

Qy 1372 ggttggccagattacacaaagtaattgattccattggttgtggccgcagacaagtcaaagcacc 1431

Qy Match

/db_xref="taxon:9606"
/clone_lib="HB0006"
/dev_stage="Adult"
/note="Organ: bocio_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

Left column (RESULT 10)

```
RESULT 10
BF331266
LOCUS       BF331266      260 bp    mRNA    linear   EST 22-NOV-2000
DEFINITION  MRO-BT0374-220300-001-h01 BT0374 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF331266
VERSION     BF331266.1   GI:11302014
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 260)
```

TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
JOURNAL   Simpson,A.J.
MEDLINE   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
          20202663
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1-6t2-MR0-BT0374-220
          300-001-f07&t3-2000-03-22&t4=1)
          Seq primer: puc 18 forward
          High quality sequence stop: 257.
FEATURES Location/Qualifiers
   source 1..257
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="BT0374"
          /dev_stage="Adult"
          /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
          SmaI; A mini-library was made by cloning products derived
          from ORESTES PCR (U.S. Letters Patent application No. 196
          /716 - Ludwig Institute for Cancer Research) profiles
          into the puC 18 vector. Reverse transcription of tissue
          mRNA and cDNA amplification were performed under low
          stringency conditions."
BASE COUNT   87 a   48 c   60 g   62 t
ORIGIN

Query Match   14.3%; Score 252.2; DB 9; Length 257;
Best Local Similarity 98.8%; Pred. No. 9.5e-44;
Matches 254; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy  1090 gattgggggagaaagaaaataacagaggaattaggaccacagtctacacagtcgaac 1149
         |||||||||||||||||||||||| |||||||||||||||| ||||||||||| |||
Db     1 GATTTGGGGGAGAAAGAAAATAACAGGAGGAATTAGGACCACAGATCTACACAGTCGGAC  60

Qy  1150 ttcaacttttattgttgaagttctgtgatgaactccaaaacaataattctaagtggag 1209
         |||||||||| || |  ||| ||||| ||||||||||| ||||||||||||||| ||||
Db    61 TTCAACTTTTATTGTTGAAGTTCTGTGTGATGAACTCCAAAACAATAATTCTAAGTGGAG 120

Qy  1210 aacttataaaggaattgtgaataatgaagaaaagtgttccaggtaactctaacttccgg 1269
         | | |||||||||||||||||||||||||||||||||| ||| |||||| |||| |||
Db   121 ACCTATAAAGGAATTGTGAATAATGAAGAAAAGTGTTTCAGGTAACTCAACTTTCGG   180

Qy  1270 gaccagtgcaaaacaacaatttcatccctcccaacctgtcgggcagatatgtcggggt 1329
         |||| |||||||||||   ||||||||||||||  |||  ||||||||||| ||||||
Db   181 GACCCAGTGCAAAACAGATTTCATCCCTCCCATCGTGCCCAGATAATGTCGGGGTTGTCCCC 240

Qy  1330 cagacatggcaccagag 1346
         |||||||||||||||||
Db   241 CAGACATGGCACCAGAG 257
```

Right column (RESULT 11)

```
RESULT 11
AA050442
LOCUS       AA050442      474 bp    mRNA    linear   EST 09-SEP-1996
DEFINITION  mj12e10.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:475914 5', mRNA sequence.
ACCESSION   AA050442
VERSION     AA050442.1   GI:1530113
KEYWORDS    EST.
```

AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          M.J. Soares,F., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
          M.J., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?t1-6t2-MR0-BT0374-
          220300-001-h01&t3-2000-03-22&t4=1)
          Seq primer: puc 18 forward.
FEATURES Location/Qualifiers
   source 1..260
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="BT0374"
          /dev_stage="Adult"
          /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
          SmaI; A mini-library was made by cloning products derived
          from ORESTES PCR (U.S. Letters Patent application No. 196
          /716 - Ludwig Institute for Cancer Research) profiles
          into the puC 18 vector. Reverse transcription of tissue
          mRNA and cDNA amplification were performed under low
          stringency conditions."
BASE COUNT   90 a   49 c   61 g   60 t
ORIGIN

Query Match   14.2%; Score 249.2; DB 10; Length 260;
Best Local Similarity 98.8%; Pred. No. 4.2e-43;
Matches 251; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy  1093 ttgggggagaaagaaaataacagaggaattaggaccacagagtctaccacagtcgaactc 1152
         ||||||||||||||||||||||| |||||||||||||||||| | ||| ||||||| |||
Db     7 TTGGGGGAGAAAGAAAATAACAGGAGGAATTAGGACCACAGACCCACACAGATCGGAACTTC  66

Qy  1153 aacttttattgttgatgaactccaaaacaataattctaagtggagacc 1212
         |||||||||| || ||||||||||||||||||||||||||| |||| |
Db    67 AACTTTTATTGTTGATGAACTCCAAAAACAATAATTCTAAGTTGGAGACC 126

Qy  1213 tataaaggaattgtgaataatgaagaaaagtgttccaggtaactctaacttccgggac 1272
         |||||||||||||||||||||||| |||||||||||| ||| |||||| |||| ||||
Db   127 TATAAAGGAATTGTGAATAATGAAGAAAAGTGTTTCAGGTAACTCAACTTTCGGGAC   186

Qy  1273 ccagtgcaaaacaatttcatccctcccatcgtgcccagatatgtcggggt 1332
         |||||||||||||   |||||||||||||||| ||||||| ||||||||
Db   187 CCAGTGCAAAACAATTTCATCCCTCCCATCGTGCCGCGTTGCTCCCCCAG 246

Qy  1333 acatggcaccagag 1346
         ||||||||||||||
Db   247 ACATGGCACCAGAG 260
```

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 474)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
  COMMENT   Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:286658
            Seq primer: -28M13 rev2 from Amersham
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1..474
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:475914"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note=Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTACAACATGTCAAGTGGGACGCGGCGGAAATTTTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
                     State Univ. from 2 1: double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia) digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo."
BASE COUNT    95 a    120 c    161 g    98 t
ORIGIN
```

```
Qy      159 cggcgcgcggcgcgactggcgcgcgcgcggggcctccggcttgctcgcggt  218
            ||||||      |||| ||   ||| ||||||| ||||||| |||| ||| |
Db      113 CGGCGGGACCATGGGGTACCGCGCGTCGTGGGCGCCGAGTGTCTGCGCCTGCTGCTGCGT  172

Qy      219 ctccgcccccggtccgcgtgcacgcgcggagggagctgggtgatggctgtgacacctagtgac  278
            | |||||||||||      |||  |  |||| ||||||||| |||  | ||  |||||| |||
Db      173 GTGTCGTCTCCGTCCGGTTCCAGCGGGAGGAGCTGGGTGCGTGTGTGGGCACATAGTGAC  232

Qy      279 ttatcaggatggtggcacaatcaataagaaattatcccgggacctcacccaatacac  338
            || ||||||| |||| ||| || |||||||||||| || |||||| || || |||||
Db      233 CTCTCAGGACACTGGGCACAATGACATCTAAGAATATTATCCCAGGGACTTACCCAAATTACAC  292

Qy      339 tgtttccgcgaaaagacaattacagaggggaaaaaacctaattctgtaggttgggga  398
            ||||| ||| ||||| |  | ||||  |||||||||| ||||||||| ||||||||||
Db      293 TGTGTGTGAAAAAGATCATCCACAGTCCCAAAAGGGGAAGAGACTTATTCTGTAGGTTGGGAGA  352

Qy      399 tttgggatatccgaaccccaaaacactgtgttctgactatctcttcaccacagctcttcaga  458
            || |||||| ||| |||||||||  ||| || || || ||| || || || || || |||
Db      399 ... 458
```

```
RESULT 12

Db      353 TTTGAACATTGAGTCCAAGAGACCTCCGCTTCGTGACTAGTCTCCTTCTTCAGCAGTGCCAACAGA  412

LOCUS       BB860022
  Qy      459 tcaaatatcgttccatactgtggaagtatgactgtcccaaaagaacctcttgttgaacacag  518
                ||||  ||||| ||| |||| | ||| || ||||| ||||||||||| |||||| |||
  Db      413 TCAGTATGGTCCATATTTGTGGAGTATGGGAGGTTCGGGCTG-TCCCAAAGAACCTCCGGCTGAACTCAAA  471

LOCUS       BB860022                 464 bp    mRNA    linear    EST 26-NOV-2001
DEFINITION  BB860022 RIKEN full-length enriched, pooled cell lines Mus musculus
            cDNA clone G43001IK05 5', mRNA sequence.
ACCESSION   BB860022
VERSION     BB860022.1  GI:17101476
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 464)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., ImoCani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,K., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            Wagi,K., Fujiwake,S., Togawa,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Inoue,K., Ishikawa,T., Ozawa,Y., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384 format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..464
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G43001IK05"
                     /clone_lib="RIKEN full-length enriched, pooled cell lines"
                     /note="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
                     (cell_line=GA3001IK05 5',
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
                     (cell_line=CRL-2116 JC), (cell_line=RCB-0545 OHTA),
```

```
Query Match             81.0%;
Best Local Similarity   13.8%;   Score 242.2;  DB 9;   Length 474;
Matches 294;  Conservative    0;   Pred. No.1.5e-41;
Mismatches 68;  Indels 1;  Gaps 1;
Db      472 CGA 474
Qy      519 tga 521
```

RESULT 13

```
LOCUS       BF086926/c
DEFINITION  RC3-HT0230-150900-112-f10 HT0230 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF086926
VERSION     BF086926.1  GI:10892636
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 451)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W., Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
            M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
  COMMENT   Contact: Simpson A.J.
```

Query Match   13.0%;   Score 228.4; DB 9; Length 464;
Best Local Similarity 80.4%;   Pred. No. 1.4e-38;
Matches   304; Conservative   0; Mismatches   71; Indels   3; Gaps   3;

```
Qy     159 cggcgcggcgcaactggccgcggccgccgctccgccggcgtcgcggt  95
           ||||| ||||||| | ||||||||||  | |||| |||| | |||||
Db     120 CGGCGGCGACCATGGGACCGGCGGGGCTGTGGCCCCAGAGG         c
Qy      88 CGGCGGGACCATGGGCAGGCGGGGCTGTGGCCCCAGAGG      147
           |||| | |||||||| |||| ||||  |||||||||||||
Db      88 CCCGGCGCCGGCTCCGGCAGGGAGGAGCTGGGTCATGGCTGCTGCGGTGTTCCCCCT  t
Qy     219 CTCCGGCGCCGGCTGCGGCAGGGAGGAGCTGGGTCATGGCTGCTGCGGTGTTCCCCCT 277
           ||||||||  ||||||| |||||||||||||||||||||
Db     148 GTGTGCTTCCGGTCCGGTTGCAGCGGAAGGAGCTGGGTCGGGTGGGGCACATAGTGAC 207
Qy     278 cttatcaggagtaggtcggcacgacatctaagaattatccggggacctacccaatcaca 337
Db     278 CTCCTCACGAACGTGGCACACGAGATATTCCAGCGACTTACCCCAACTACA  267
Qy     338 ctgttcgcaaagacaattacagtaccaaaaggtggaaaagaactgattctgaggtgggag 397
Db     208 CCTCTCAGCAGACGGTGGCACACGAGATATTCCAGCGACTTACCCCAACTACA  267
Qy     398 atttggtatatcggataccagaacgactgtgctctctgactatatctctcctttca-ccagctcttca   456
Db     268 CTGTGTCGTGAAAAGACATCACACGCCTCCAAAAGGGGAGAGACAGACTTAATTCTGAGGTTGGGAG 327
Qy     457 gatcaatatgtccacactctggaagtatgactgttcccaagaaaactcctgtgtgaacaca   516
Db     388 GATCACAGTATGGTCCATATATTGTGGGAGTTGGGCTG-TCCCAAAGAACTCCGGCTAGAACTCA  446
Qy     517 agtggaagtaacctccgc  534
Db     447 AACGAACTGCACTGTCCTC  464
```

RESULT 14

```
LOCUS       AA042037
DEFINITION  mj04N13.r1 Soares mouse embryo NbME13.5 Mus musculus cDNA
            clone IMAGE:475175 5', mRNA sequence.
ACCESSION   AA042037
VERSION     AA042037.1  GI:1520227
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
  COMMENT   Contact: Marra M/Mouse EST Project
```

Query Match   12.0%;   Score 212.2; DB 10; Length 451;
Best Local Similarity 96.4%;   Pred. No. 3.9e-35;
Matches   217; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

```
Qy    1535 tccttgttgctcctggtgttgctggcaggaagtggggggatctgagcccttagaagaaggaggaga 1594
           ||||||||||||||||||||||||||||| |||||||| |||||||||||||||||||||||||||
Db     449 TCGACGGTGTTGCTGGTGTTTGCTGGAATGGGGAGTCTTTCAGCCCTTTAGAAGAAGGAGGAGA  390
Qy    1595 agaaagaagtccgtatggggcccagaaaaacagactgttgggacgagatca 1654
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     389 AGAAAGGAAGTCCGTATGGGGATCAGCAGGAGGCTCAGAAAAACAGACTGTTGGGACGAGATTA  330
Qy    1655 aaatatcccttgccagacatcagtcagctgagttcaccatcaggatatgataggacaggagg 1714
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     329 AAATATCCCTTTGCCAGACATCAGTCAGCTGAGTTCACCATCAGGTATGATAATGAGGAAGG  270
Qy    1715 agatgacacaaaagttagatctcatcacacagttaggcaggtt 1759
           ||||||||||||||||||||||||||||||||||||||||||||
Db     269 AGATGACACAAAAGTTAGATTCTCATCACACAGTCGATATGGCAGCTT 225
```

```
ORIGIN
BASE COUNT   104 a   125 c   110 g   112 t
```

```
ORIGIN
BASE COUNT   95 a   120 c   153 g   96 t
```

(cell_line=RCB-0559 K-1, F1), (cell_line=RCB-1283 B16
melanoma), (cell_type=B cells), (cell_line=CRL-1702 WEHI 231
), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
(cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
(tissue_type=bladder, cell_line=RCB-0554 MBT-2),
(tissue_type=bone marrow, cell_type=stroma cell,
(tissue_type=CRL-2028 SR-4987), (tissue_type=kidney,
cell_line=CRL-0549 Cle-H3), (tissue_type=submandibular gland,
cell_line=CCL-142 RAG), (tissue_type=colon,
cell_line=1734 SCA-9, clone 15; (strain=BALB/C/
cell_line=1734 RAG), (strain=C3H, clone 13.20-3B3
cell_type=B cells, cell_line=1669 BCl1 clone 13.20-3B3
), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
BC3H1)*

Washu-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousees@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:285519

Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 399.

FEATURES   Location/Qualifiers
source   1..406
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:475175"
            /clone_lib="Soares mouse embryo NbME13.5 14.5"
            /sex="unknown"
            /tissue_type="embryo"
            /dev_stage="13.5-14.5dpc total fetus"
            /lab_host="DH10B"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5'
            TGTTACCAATCTGAAGTGGCAGCGCCCCGGAAATTTTTTTTTTTTTTTTTTTTT
            T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
            14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
            State Univ., from 2]; double-stranded cDNA was ligated to
            Eco RI adaptors (Pharmacia), digested with Not I and
            cloned into the Not I and Eco RI sites of the modified
            pT7T3 vector. Library went through one round of
            normalization, and was constructed by Bento Soares and
            M.Fatima Bonaldo.

BASE COUNT   75 a   109 c   141 g   81 t
ORIGIN

Query Match   11.9%; Score 209.4; DB 9; Length 406;
Best Local Similarity   82.5%; Pred. No. 1.5e-34;
Matches 240; Conservative   0; Mismatches   51; Indels   0; Gaps   0;

Qy   159 cgggcgcggcgcaactggcgcggggcggcggggcctcctggcttgctccggt 218
           |||||| ||||||||||||||||||||||||||||||||||||||| |||||
Db   116 CGGCGGGACCATGTCGACCGGGGCTGGTGGGGCCGAGTGTCCTGCGCGCGCCGT 175

Qy   219 ctccgccccggcccggctcggcgaggagggcctggtgatgctggaacctagtgac 278
           ||||| ||||| |||||||||||||||||| |||||| || ||||||||||||||
Db   176 GTGTGCTCCGGCTCCGGCTTGCAGCGCGAGCAACCTGGCGTGATGGCGCTGTGGCGCACATAGTGAC 235

Qy   279 ttatcaggatagtggcacaatctaagagaatatcccgggacctacccaatcacac 338
           ||||||| |||| |||||||| ||||||||||||||||||||||||||||||||
Db   236 CTCTCACGACAGTGGCACAATGCATCTAAGAAATTANTCCAGGCAGCGACTTACCCCAAYTACAC 295

Qy   339 tgtttgcgaaagacaattacagtaccaaaggggaaaggaactgagttctggggaga 398
           |||||||||||| ||||||||||||||||||||||| ||||||||||||||||||
Db   296 TGTGTGTGAAAAGAATCATCACACATCCCAAAACGCGGAAGAGACTTATTCTGAGGTTGGCGAGA 355

Qy   399 tttggatatccggaatcccagacccgtgcttctgactatctctctccaccag 449
           ||||||||| |||||||||||||||| ||||||||||||||||||||||||
Db   356 TTTGAACATTGAGTCCAAGACCTGCGCTTCTGACTATCTCCTCTTTCAGCAG 406

REFERENCE   1 (bases 1 to 896)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
            NIH-MGC http://mgc.nci.nih.gov/,
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: NCI-CGAP clone distribution Information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9255 row: i column: 21
            High quality sequence stop: 628.

FEATURES   Location/Qualifiers
source   1..896
            /organism="Mus musculus"
            /strain="CZECH II"
            /db_xref="taxon:10090"
            /clone="IMAGE:4013276"
            /clone_lib="NCI_CGAP_Lu30"
            /tissue_type="tumor, metastatic to mammary"
            /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; transgenic model WNT-1, expression driven by
            MMTV-LTR enhancer; cloned unidirectionally. Primer: Oligo
            dT: library constructed by Life Technologies.
            Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT   229 a   228 c   248 g   191 t
ORIGIN

Query Match   11.5%; Score 203.2; DB 10; Length 896;
Best Local Similarity   84.2%; Pred. No. 3.9e-33;
Matches 229; Conservative   0; Mismatches   43; Indels   0; Gaps   0;

Qy   1489 acatccacaggaataaacattacaacggtggctattccattggtgctcctggtgtcctg 1548
           |||||||||||||| |||||||| |||| ||||| ||||||||| |||||||||||||
Db   29 ACAACCCAGAAGANTGAAGAATATTACAACTGTGGCGGATTCCATCACGATCTTCANCCGCCCTC 88

Qy   1549 gtgtttgctggaattgggggatcttcagccttagaaaggaagaaggaagagaagtccg 1608
           ||||||| ||||| |||||||||||| ||||||||||||||||||||||||||||||
Db   89 CTTCTGACTGGAATGGGGGCATGCTTTGCANTCTGTAGAAGAGAGGAAAAGAAGAAAAATCCA 148

Qy   1609 tatggatcagcgaggcctcaagaaaaaacaagacgtgtggaagcagaatcccttggcc 1668
           ||||| ||||||||||||| |||||||||| |||||||||||| |||||||||||||
Db   149 TATGTCGTCAGCTGAACGCCTCACAAAAAACAGGCGTGTGGAANGCAGAGATTNAAAATATCCCTTGCC 208

Qy   1669 agacatcagtcaactgagttttaccatcagcctatataatagagagagggagtgacacaaaag 1728
           || |||||| ||||||||||||||||||||||||||| |||||||||||||||||||||
Db   209 AGGCATCACGTCGACGGAATTACCATCACCATATCAATAGAAATGAAAANGACAGACACAAAAG 268

Qy   1729 ttagatctccatccacaagtgatatgggagagtta 1760
           ||||||| |||||| |||||||||| |||||||||
Db   269 TTGGATCTCATCACTAGTGATAATGGCCAGATTA 300

Search completed: July 9, 2002, 00:10:02
Job time: 2629 sec

Total number of hits satisfying chosen parameters: 767066

Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Searched: 383553 seqs, 12281672 residues

Scoring table: IDENTITY_NUC
                         Gapop 10.0 , Gapext 1.0

Sequence: 1 atggatccggtgcagggca.........caagtgatggcaggttaa 1761

Perfect score: 1761
Title:

Run on: July 8, 2002, 23:34:13 ; Search time 78.07 Seconds
                                                  5540.676 Million cell updates/sec
                                                  (without alignments)

OM nucleic - nucleic search, using sw model

                         Copyright (c) 1993 - 2000 Compugen Ltd.

                         GenCore version 4.5

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA.*
                    1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                    2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                    3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                    4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                    5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                    6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	92.8	5.3	6909	2	US-08-804-196-1	Sequence 1, Appli
2	92.8	5.3	6909	9	US-09-659-235C-9	Sequence 9, Appli
3	92.8	5.3	6909	3	US-08-746-111-26	Sequence 1, Appli
4	81.6	4.6	7032	4	US-07-607-538C-1	Sequence 3, Appli
5	80.6	4.6	6585	3	US-09-324-867-1	Sequence 26, Appli
6	78.4	4.5	4629	2	US-08-746-111-4	Sequence 4, Appli
7	78.4	4.5	4670	3	US-08-488-891-7	Sequence 1, Appli
8	78.4	4.5	4999	4	US-08-717-294-41	Sequence 41, Appli
9	78.4	4.5	4999	4	US-08-717-618-14	Sequence 14, Appli
10	78.4	4.5	4999	2	US-09-364-862-14	Sequence 14, Appli
11	78.4	4.5	5035	2	US-08-882-083-1	Sequence 1, Appli
12	78.4	4.5	5035	3	US-09-243-539-1	Sequence 1, Appli
13	78.4	4.5	5035	3	US-08-558-107-1	Sequence 1, Appli
14	78.4	4.5	6999	1	US-08-276-594A-1	Sequence 1, Appli
15	78.4	4.5	7056	1	US-09-121-202-1	Sequence 1, Appli
16	78.4	4.5	8241	6	5171844-1	Patent No. 5171844
17	78.4	4.5	8967	1	US-08-336-851A-1	Sequence 1, Appli
18	78.4	4.5	9009	3	US-07-864-004B-3	Sequence 3, Appli
19	73.4	4.5	9009	3	US-08-251-937A-3	Sequence 3, Appli
20	78.4	4.5	9009	1	US-08-212-133A-1	Sequence 1, Appli
21	78.4	4.5	9009	2	US-08-670-707A-1	Sequence 1, Appli
22	78.4	4.5	9009	1	US-08-670-707A-1	Sequence 1, Appli
23	78.4	4.5	9009	1	US-09-037-601-3	Sequence 3, Appli
24	78.4	4.5	9009	1	PCT-US94-13200-1	Sequence 1, Appli
25	78.4	4.5	9354	1	US-08-683-839B-2	Sequence 2, Appli
26	78.4	4.5	11933	1	US-09-470-618-13	Sequence 13, Appli
27	78.4	4.5	11993	4	US-09-364-862-13	Sequence 13, Appli
28	58	3.3	2303	2	US-08-480-229C-9	Sequence 9, Appli
29	58	3.3	2303	2	US-09-659-235C-9	Sequence 9, Appli
30	57.6	3.3	1384	1	US-07-607-538C-1	Sequence 1, Appli
31	57.6	3.3	1384	1	US-08-162-402B-1	Sequence 1, Appli
32	57.6	3.3	1934	2	US-08-162-402B-7	Sequence 7, Appli
33	54.4	3.1	4451	3	US-08-717-294-42	Sequence 42, Appli
34	53.8	3.1	318	2	US-08-480-229C-19	Sequence 19, Appli
35	53.8	3.0	318	2	US-09-659-235C-19	Sequence 19, Appli
36	53.4	3.0	7493	4	US-09-037-601-5	Sequence 5, Appli
37	53.4	3.0	7493	1	US-08-474-503-5	Sequence 5, Appli
38	53.4	3.0	7493	2	US-08-670-707A-38	Sequence 38, Appli
39	53.4	3.0	7493	1	US-09-037-601-38	Sequence 38, Appli
40	53.4	3.0	4334	2	US-08-670-707A-36	Sequence 36, Appli
41	52.8	3.0	4334	4	US-08-670-707A-36	Sequence 36, Appli
42	52.8	3.0	6402	2	US-09-037-601-36	Sequence 36, Appli
43	52.8	3.0	6402	4	US-09-037-601-36	Sequence 36, Appli
44	52.8	3.0	6402	4	US-08-480-229C-11	Sequence 11, Appli
45	52.2	3.0	1780	2	US-08-480-229C-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-08-804-196-1

Sequence 1, Application US/08804196

; Patent No. 5874256
; GENERAL INFORMATION:
; APPLICANT: Reitsma, Rogier
; APPLICANT: Bertina, Pieter
; TITLE OF INVENTION: A method for diagnosing an increased
; TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
; TITLE OF INVENTION: thrombosis and kit for use with the same.
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5874256e1 Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,196
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/454,353
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 948-7400
; TELEFAX: (301) 948-9751
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA

US-08-804-196-1

Query Match   5.3%; Score 92.8; DB 2; Length 6909;
Best Local Similarity 51.0%;
Matches 248; Conservative 0; Mismatches 232; Indels 6; Gaps 2;

Qy   898 aatggttgcagcagcatcctgagtttgaacctgacgacggcaaatcagagcttcttcctca 957

```
Db  6280  AATGGATGGTTCCACACCCTGGGTATGGAAAAATGAAAAATGAAAGAATGAAAAGCAAAATCACA  6339
          |||| |                              |||              |||
Qy   958  tggcagtcggtcaatgagagtggagagacaagttcactggtctccggccaagcccgacct  1017
          |||  ||| || |  ||| |  ||||||   ||  || ||  |||      |||
Db  6340  GCTTCTTCGTTTAAGAAAATCTTGGTGGGGAGAGATTACTGGGAACCCTTCCGTGCCGTCTG  6399
          | |||  |||                      ||| |||
Qy  1018  caggaccaaaggcccatcatgggcttcgggcgcgacagtagcaacaacaacacacgagag  1077
          | ||| || |||                    ||||
Db  6400  AATGGCCCAGGGACGTGTGAATGCCTGGCACAAGCCCAAGCAAACAACAACAATAAGC------AG  6453
          ||||                            |||                        ||
Qy  1078  tggctggggagtcggatcgggggacagaaaataaccaggaaatctaggaccacaggacct  1137
          ||| | ||| |||                   || ||| |||
Db  6454  TGGCTAGAAAATTGATTCTACTCAAGATCAAGAAGAAATACGGCCAATTATAACACAGGGCCTGC  6513
          |||||                          |||  ||| |||
Qy  1138  acacagtcgaactctcaacttctatgtcagatagagtttttgtgaactttcaaaacacat  1197
          | |||  ||| |||                  ||  ||| |||
Db  6514  AAGTCTGTCTCCTGGTCGAAAAGTATGTATAAAGGACTATACCATCCACTACAGTGCAGACGGA  6573
          |||  ||| |||                    ||||||
Qy  1198  tctaagtggaagacctataaggaattgtgaataatgaagaaaagtgttttcaggtaac  1257
          ||| ||| |||                     ||||
Db  6574  GTGGAATGGAAAACCATACAGGCTGAAATCCTCCATGGTGGACAAGAGCTTGGCCTT  6633
          ||| |||                         ||||||
Qy  1258  tctaacttcgggaccagtgcaaaacaattcatccctcccatcctggccgggatatgtg  1317
          ||||                            ||||||
Db  6634  ACTAATACCAAAGGACACATGTGAAAGAACTTTTTCAACCCCCCAATCATTCCAGGTTTATC  6693
          |||||                           |||||
Qy  1318  cgggttgtccccagacactggcaccaggagatagcctgaagtggggcatgttga  1377
          |||                             |||
Db  6694  CGTGTCATTCCTAAAACATGGAATCAAAGTATTACACTTCGCCTGGAACTCTTTGGCTGT  6753
          ||||                            ||||
Qy  1378  cagatt 1383
          | |||
Db  6754  GATATT 6759
```

```
RESULT 2
US-08-658-340-1
Sequence 1, Application US/08658340
Patent No. 5910576 5861489
Patent No. 5910576 5861489el Patent Department

GENERAL INFORMATION:
   APPLICANT: Bertina, Rogier
   APPLICANT: Reitsma, Pieter
   TITLE OF INVENTION: A method for diagnosing an increased
   TITLE OF INVENTION: risk for thrombosis or a genetic defect causing
   TITLE OF INVENTION: thrombosis and kit for use with the same.
   NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Akzo No. 5910576 5861489el Patent Department
   STREET: 1300 Piccard Drive, Suite 206
   CITY: Rockville
   STATE: Maryland
   COUNTRY: USA
   ZIP: 20850
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/658,340
   FILING DATE:
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 08/454,353
   FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
   NAME: Gormley, Mary E
   REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (301) 948-7400
   TELEFAX: (301) 948-9751
INFORMATION FOR SEQ ID NO: 1:
```

```
SEQUENCE CHARACTERISTICS:
   LENGTH: 6909 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: both
   TOPOLOGY: unknown
   MOLECULE TYPE: cDNA
US-08-658-340-1

Query Match        5.3%;    Score 92.8;  DB 2;  Length 6909;
Best Local Similarity  51.0%;  Pred. No. 1.2e-13;
Matches  248; Conservative  0; Mismatches  232;  Indels  6;  Gaps  1;
```

```
Qy   898  aatggttgcaagacagatcctgagtttgaacctgacgagcaaatcagagcttcttcctca  957
          |||| |                              |||              |||
Db  6280  AATGGATGGTTCCACACCCTGGGTATGGAAAAATGAAAAATGAAAGAATGAAAAGCAAAATCACA  6339
          |||  ||| || |  ||| |  ||||||   ||  || ||  |||      |||
Qy   958  tggcagtcggtcaatgagagtggagagacaagttcactggtctccggccaagcccgacct  1017
          | |||  |||                      ||| |||
Db  6340  GCTTCTTCGTTTAAGAAAATCTTGGTGGGGAGAGATTACTGGGAACCCTTCCGTGCCGTCTG  6399
          | ||| || |||                    ||||
Qy  1018  caggaccaaaggcccatcatgggcttcgggcgcgacagtagcaacaacaacacacgagag  1077
          ||||                            |||
Db  6400  AATGGCCCAGGGACGTGTGAATGCCTGGCACAAGCCCAAGCAAACAACAACAATAAGC------AG  6453
          ||| | ||| |||                   || ||| |||
Qy  1078  tggctggggagtcggatcgggggacagaaaataaccaggaaatctaggaccacaggacct  1137
          |||||                          |||  ||| |||
Db  6454  TGGCTAGAAAATTGATTCTACTCAAGATCAAGAAGAAATACGGCCAATTATAACACAGGGCCTGC  6513
          | |||  ||| |||                  ||  ||| |||
Qy  1138  acacagtcgaactctcaacttctatgtcagatagagtttttgtgaactttcaaaacacat  1197
          |||  ||| |||                    ||||||
Qy  1198  tctaagtggaagacctataaggaattgtgaataatgaagaaaagtgttttcaggtaac  1257
          |||  ||| |||                    ||||||
Db  6514  AAGTCTGTCTCCTGGTCGAAAAGTATGTATAAAGGACTATACCATCCACTACAGTGCAGACGGA  6573
          ||||                            ||||||
Qy  1258  tctaacttcgggaccagtgcaaaacaattcatccctcccatcctggccgggatatgtg  1317
          |||                             |||||
Db  6574  GTGGAATGGAAAACCATACAGGCTGAAATCCTCCATGGTGGACAAGAGCTTGGCCTT  6633
          ||||                            ||||
Qy  1318  cgggttgtccccagacactggcaccaggagatagcctgaagtggggcatgttga  1377
          |||||                           ||||
Db  6634  ACTAATACCAAAGGACACATGTGAAAGAACTTTTTCAACCCCCCAATCATTCCAGGTTTATC  6693
          | |||                           ||||
Qy  1078  tggctggggagtcggatcgggggacagaaaataaccaggaaatctaggaccacaggacct  1137
          | |||                           ||||
Db  6400  AATGGCCCAGGGACGTGTGAATGCCTGGCACAAGCCCAAGCAAACAACAACAATAAGC------AG  6453
          ||||                            ||||
Db  6454  TGGCTAGAAAATTGATTCTACTCAAGATCAAGAAGAAATACGGCCAATTATAACACAGGGCCTGC  6513
          |||                             ||||
Db  6514  AAGTCTGTCTCCTGGTCGAAAAGTATGTATAAAGGACTATACCATCCACTACAGTGCAGACGGA  6573
          ||||                            ||||
Qy  1138  acacagtcgaactctcaacttctatgtcagatagagtttttgtgaactttcaaaacacat  1197
          | |||                           ||||
Db  6574  GTGGAATGGAAAACCATACAGGCTGAAATCCTCCATGGTGGACAAGAGCTTGGCCTT  6633
          ||||                            ||||
Qy  1198  tctaagtggaagacctataaggaattgtgaataatgaagaaaagtgttttcaggtaac  1257
          ||||                            ||||
Db  6634  ACTAATACCAAAGGACACATGTGAAAGAACTTTTTCAACCCCCCAATCATTCCAGGTTTATC  6693
          ||||                            ||||
Qy  1318  cgggttgtccccagacactggcaccaggagatagcctgaagtggggcatgttga  1377
          ||||                            ||||
Db  6694  CGTGTCATTCCTAAAACATGGAATCAAAGTATTACACTTCGCCTGGAACTCTTTGGCTGT  6753
          ||||                            ||||
Qy  1378  cagatt 1383
          | |||
Db  6754  GATATT 6759
```

```
RESULT 3
US-08-746-111-26
Sequence 26, Application US/08746111
Patent No. 6066778

GENERAL INFORMATION:
   APPLICANT: Ginsburg, David
   APPLICANT: Cui, Jisong
   TITLE OF INVENTION: Compositions And Methods For Screening
   TITLE OF INVENTION: Compounds For Anticoagulant Activity
   NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Medlen & Carroll, LLP
   STREET: 220 Montgomery Street, Suite 2200
   CITY: San Francisco
   STATE: California
   COUNTRY: United States of America
   ZIP: 94104
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/746,111
FILING DATE: 06-NOV-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
  NAME: Ingola, Diane E.
  REGISTRATION NUMBER: 40,027
  REFERENCE/DOCKET NUMBER: UM-02536
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 705-8410
  TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
  LENGTH: 6909 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
  MOLECULE TYPE: cDNA

US-08-746-111-26
```

Query Match 5.3%; Score 92.8; DB 3; Length 6909;
Best Local Similarity 51.0%; Pred. No. 1.2e-13;
Matches 248; Conservative 0; Mismatches 232; Indels 6; Gaps 1;

```
Qy     898  aatggttgcagcagtacttctggtttgaacctgatcgcaggcaaatcagagctctctctca    957
             || ||| |||||| | |  ||| ||| ||  ||||| || ||| ||||| ||| ||| |
Db    6280  AATGGAGATTCACACCCCTGGGTATGGAAAAATGGAAAAGAAAGGAAGCAAAAACAAAATCACA  6339

Qy     958  tggcagtcggtcaatgagtggagacaagttcactggtctcctggccaagcccgactt  1017
             ||||| || ||||||  | |||  ||| ||  ||||| || |||| ||| ||| |||
Db    6340  GCTTCTTCGTTTAAGAAAATTCTTGGTGGGGAGAGTTACTGGGGAAACCTTCCGTGCCCGTTCTG  6399

Qy    1018  caggacacaaggcccatcatggggtcttcaggcgacagtagcacacaacaacaaccacgaagg  1077
             ||||||||||| ||||| ||| ||| ||  |||| | ||  || |||| | || | ||||
Db    6400  AATGCCCAGGGACGGTCGTGAATGCCTGGCAAGGCCAAACACAACAATAAGC------AG  6453

Qy    1078  tggctgagagatcgatcttgggggagaaaaataacaagaaaattaggagccacagagtct  1137
             |||||| || ||||  ||||||| |||||||| || ||||||| | |||| |||||| |||
Db    6454  TGCCTAGAAAATGACTACTACTCAGAGATCAAGAAGAGATAAACGGAGTAAGAAGAGCGGCTCC  6513

Qy    1138  acacagtcggaacttcaacttttatgttaagagtttgtgatgaacttcaaaaacaatat  1197
             || ||| | |||| | ||||||||  | |||||||| |||| |||| || |||| |||||
Db    6514  AAGTCTCTGTCCTTCTGAAAAATGTATGGTAAAGAGCTATACCACTACAGTCAGTCAGCAGGGA  6573

Qy    1198  tctaagtggagaagacctatcaaaggaattgtgaatattgaagaaaaggtgtttcaaggtaac  1257
             |||||| | || || || ||||| ||| ||||||| || |||||||||| | | |||| ||
Db    6574  CTGGAATGGAAAACCATACAGGCTGAAATCCTCCATGGTGGACACAGATTTTTGAAGGAAT  6633

Qy    1258  tctaacttcggggaccagtgcaaaacaattcatcctcctccatcgtggccagatatgtg  1317
             ||| ||| |||| || ||| | |||||| ||  ||||||||||| ||| |||||  |||
Db    6634  ACTAATACCAAAAGGACATGTGAGAAGAACCACTTTTTCAACCCCCAAATCATTATCCAGGTTTATC  6693

Qy    1318  cgggttgtgtcccccagacatggcccccagaggatagcctcgaaggtggagctcattggttgc  1377
             || ||||||| |||| || ||| | | ||| ||||| || |||| || ||||| ||| |||
Db    6694  CGTGTCATTCCTAAAACATGGAATCAAAGTATTACACTTCGCCGTGGGAACTCTTTGGCTGT  6753

Qy    1378  cagatt  1383
             ||| ||
Db    6754  GATATT  6759
```

```
RESULT 4
US-09-324-867-1
Sequence 1, Application US/09324867A
Patent No. 6251632

GENERAL INFORMATION:
APPLICANT: Lillicrap, David
APPLICANT: Cameron, Cherie
APPLICANT: No. 6251632ley, Colleen
APPLICANT: Horrocks, L. Suzanne Hoyle
APPLICANT: Hough, Christine
TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
FILE REFERENCE: 1669.0010002/JAG/BJD
```

```
CURRENT APPLICATION NUMBER: US/09/324,867A
CURRENT FILING DATE: 1999-06-03
EARLIER APPLICATION NUMBER: 09/035,141
EARLIER FILING DATE: 1998-03-03
EARLIER APPLICATION NUMBER: 09/035,141
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/039,953
EARLIER FILING DATE: 1997-03-06
NUMBER OF SEQ ID NOS: 63
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
  LENGTH: 7032
  TYPE: DNA
  ORGANISM: Canis familiaris
  FEATURE:
  NAME/KEY: CDS
  LOCATION: 1..7029
US-09-324-867-1
```

Query Match 4.6%; Score 81.6; DB 4; Length 7032;
Best Local Similarity 54.6%; Pred. No. 7.5e-11;
Matches 212; Conservative 0; Mismatches 164; Indels 12; Gaps 2;

```
Qy     994  tggtcctggccagaccaagagcccatcatgggctgcggccggcggaccagt  1053
             ||| ||  ||| ||||| ||||||| ||||| | | ||||| ||||||||||||
Db    6640  tgggtcctctcccaagccgcctgcaccctgaccggggc------aggacaatgctgcgggagga  6693

Qy    1054  agcacacaacaacaaccacagagctgggaactcaggtcaaagttggctggggaatggaaaaata  1113
             || | | ||||| ||| |||||| | ||||  ||| |||||  |||||| ||| |||| |||
Db    6694  cctcaggcaaataaacccaaaagagtggtctgcaagtcaaggttggggagagaaaaagaaaata  6753

Qy    1114  acaggaataacacacaagggggtgaaatcctccatcagcagcatgtatgtcaagaggttc  1173
             || ||||| ||| | |||| ||  |||||||| ||| || |  ||| ||| ||||| ||
Db    6754  acaggaataacacacaggggtgaaatcctccatcagcatgtatgtcaagaggttc  6813

Qy    1174  gtgatgaacttcaaaaacaatattccttcaagtggaagaacctataaaggaattgtgaataat  1233
             ||||| ||||| ||||||||| | || ||| || ||||| ||| ||||| || |||||| |
Db    6814  ctcatcccagtagtcaagtgccatcaactaactggactctgttc------ttcagatagc  6867

Qy    1234  gaagaaaaggtgtttcaaggttagctacctttgggaccctgccaaacaacttcatc  1293
             ||| ||||| | | |||||| || ||  ||||||| ||| | ||| ||| ||||
Db    6868  aagtcaaggtctctccaggaaaccggaactcctccacgcctgtgggaaccgtctga  6927

Qy    1294  cctccactcgtgccagatatgtgcgggttgtgtcccccagacatggcccccagaggatatgcc  1353
             || ||| |||| ||||||||||| ||| |||| ||||| || |||| | ||||| |||| ||
Db    6928  ccccgctggtggctcgctacctggtgccggagctgcaccgcagacatcgcc  6987

Qy    1354  ttgaagtggagctcattggttgccaga  1381
             ||||| |||||| ||||| | |||||
Db    6988  ctgagcctggaaggcctggggctgcgaca  7015
```

```
RESULT 5
US-08-746-111-4
Sequence 4, Application US/08746111
Patent No. 6066778
GENERAL INFORMATION:
APPLICANT: Ginsburg, David
APPLICANT: Cui, Jisong
TITLE OF INVENTION: Compositions And Methods For Screening
TITLE OF INVENTION: Compounds For Anticoagulant Activity
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Medlen & Carroll, LLP
  STREET: 220 Montgomery Street, Suite 2200
  CITY: San Francisco
  STATE: California
  COUNTRY: United States of America
  ZIP: 94104
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
```

US-08-746-111-4

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/746,111
  FILING DATE: 06-NOV-1996
  CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
  NAME: Ingolia, Diane E.
  REGISTRATION NUMBER: 40,027
  REFERENCE/DOCKET NUMBER: UM-02536
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 705-8410
  TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
  LENGTH: 5585 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
  DESCRIPTION: /desc = "DNA"
FEATURE:
  NAME/KEY: CDS
  LOCATION: 6..6554

Query Match   4.6%;   Score 80.6;   DB 3;   Length 6585;
Best Local Similarity 52.1%; Pred. No. 1.3e-10;
Matches 207; Conservative   0; Mismatches 184; Indels   6; Gaps   1;

Qy   987 agttcactggtctcctggccaaggcccgacttcaaggacacaaggacccatggggttcggg 1046
          ||||| ||| || || | ||| || ||| || ||| | |||| ||||| | |||| ||
Db   6161 AGACTACTGGGAGCCCTCCTTGCCCGCTGCCAGGCCCCGGCCCCGCGTCACCGCCTGGCA 6220

Qy   1047 cgacagtagcacaccacaacacacagaggtgctggaagtcgattgttgggggagaaaa 1106
          | || | ||| | || ||| |||| |||| | | || |
Db   6221 AGCCAAGGCAAACAACAACAACAGC-------AGTGGTTACACAGTCGATCGTCAAAATCAA 6274

Qy   1107 gaaaatacacagaggaattaggacccacaggatctacacagtcgaacttcaactttatgttaa 1166
          | | | | |||| ||| ||| || | || | || | || |||| | ||| ||
Db   6275 GAAGGTAACGGCCATCGTAACCACACGGCGCTGTAAGTCTCGTCCTCTGAGATGTACCTGAA 6334

Qy   1167 gagttcttgatgaactccaaaacaaccaataacttctaagtcgagagaacctataaggaattgt 1226
          | || || | || || || || ||| || | || | ||| || | || | ||||
Db   6335 GAGCTACAGCATCCAGTACNGTACAGCACACGGGCTGTGGCCACNGGAAACCTTACCGACAGAAATC 6394

Qy   1227 gaataatgaagaaaagtgttcagggtaactctaactttcgggaccagtgcaaacaa 1286
          | || ||| | | ||| || | ||| | | ||| || | ||||| |||||
Db   6395 CTCCATGGTGGACAAGATTTTGAGGAAACAGCAATACCAAGGCGACAGAGAACACTT 6454

Qy   1287 tttcatcctcctccatcgtggccagatatgtgcgggttgtccccagacatggccacagag 1346
          |||| | || || || || |||| | | | ||| | ||| | ||||| | |||
Db   6455 TTTCAACCCGCCCCAATTATTTCCGACATCCATCCNTTTCAAAAACATGGAACCAGCAGAG 6514

Qy   1347 gatagcctctgaaggtgggagctcattggttccgaagatt 1383
          | || | ||| ||| || || | || || | | |
Db   6515 CATCGCCCTTCGCCTAGAGCTCTTCGCCTGTGACATT 6551

RESULT 6
US-08-484-891-7
Sequence 7, Application US/08484891
Patent No. 5935935
GENERAL INFORMATION:
  APPLICANT: Connelly, Sheila
  APPLICANT: Kaleko, Michael
  APPLICANT: Smith, Theodore
  TITLE OF INVENTION: Adenoviral Vectors for
  TITLE OF INVENTION: Treatment of Hemophilia
  NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
  ADDRESSEE: Cecchi, Stewart & Olstein
  STREET: 6 Becker Farm Road

CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
  MEDIUM TYPE: 3.5 inch diskette
  COMPUTER: IBM PS/2
  OPERATING SYSTEM: MS-DOS
  SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/484,891
  FILING DATE: 07-JUN-1995
  CLASSIFICATION: 514
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/218,335
  FILING DATE: 25-MAR-1994
  APPLICATION NUMBER: 08/074,920
  FILING DATE: 10-JUN-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Olstein, Elliot M.
  REGISTRATION NUMBER: 24,025
  REFERENCE/DOCKET NUMBER: 271010-273
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 201-994-1700
  TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
  LENGTH: 4629 bases
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
MOLECULE TYPE: cDNA primer
FEATURE:
  NAME/KEY: Factor VIII cDNA with
  NAME/KEY: B domain deleted
US-08-484-891-7

Query Match   4.5%;   Score 78.4;   DB 2;   Length 4629;
Best Local Similarity 54.1%; Pred. No. 4e-10;
Matches 210; Conservative   0; Mismatches 166; Indels 12; Gaps   2;

Qy   993 ctggtctcctggccaaggcccgacttcaaggacacaaggacccatggggcttcggggcgacag 1052
          |||||| |||||||| ||| ||||| |||| | | || |||||||| || ||||| |
Db   4023 CTGGTCTCCTTCAAAAGGCTGGACTCACCTTCACCTCCAAGG-----GAGGAGTAATGCCTGGAG 4076

Qy   1053 tagcacacaaccacaacacacagaggtgctggaagtcgattgttggggggagaaaagaaaat 1112
          || | ||| ||||| | || ||| || |||| ||| ||||| || || | ||
Db   4077 ACCTCAGCGGTGAATAATCCANAAAGNGTGGCCTGCCAGGCCTGGAGGCCGCAAGGGAGAAGAT 4136

Qy   1113 aacagagaattaggaccacagaggatctacacagtcgaacttcaactttatgttaagagtt 1172
          ||||| | | ||| |||| || | || | ||| || |||| || | ||||||
Db   4137 CACAGAGACTAACTACTCAGGGAGTAAAATCTCGTTACCACCATGTATGTCGAACGAGAGTT 4196

Qy   1173 tgtgatgaactccaaaacaacaataattctaagtggaggacctataaggaattgtgaattcat 1232
          ||| | || |||| |||| |||| | | || | ||| | | | | |||| |||
Db   4197 CCTCANCTCCAGCAGTCAAGATGGCCATCAGTCAGTGGCACTCTCTTT------TTCACAAATGG 4250

Qy   1233 tgaagaaaaggtgttcagggtaactctaacttcgggaccagtgcaaacaa 1292
          || || || | || | || ||| | | || | || | |||| |||||
Db   4251 CAAAGTAAAGGTTTTTCAGGGAAATCAAGACTCGTTCACACCTCGTGCTGAGTGGTGGAACTCTCTAGA 4310

Qy   1293 ccctccaccgtggccagatatgtgcgggttgtccccagacatggccacag 1352
          |||||| | || ||||| | | || | || | ||| | |||||||
Db   4311 CCCACCGTTACTGACTCCGTACCTTCGAATTCACCCCCAGAGTTGGGTGCGACACGAGATTGC 4370

Qy   1353 cttgaagttgggagctcattggttgccag 1380
          |||| || || || | || || |||
Db   4371 CCTGAGGATGAGAGGTTCTGGGCCTGCCAG 4398

RESULT 7
US-08-717-294-41

Sequence 41, Application US/08717294
Patent No. 6111148
GENERAL INFORMATION:
TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF
APPLICANT: SEED, BRIAN
APPLICANT: HAAS, JURGEN
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/717,294
FILING DATE: 20-SEP-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/345001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 4670 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-717-294-41

Query Match   4.5%;   Score 78.4;   DB 3;   Length 4670;
Best Local Similarity   54.1%;   Pred. No. 4e-10;
Matches 210;   Conservative   0;   Mismatches 166;   Indels 12;   Gaps 2;

Qy   993 ctggtctcctggccaagccgacttcaggaccaagcgccatcggcttcggcgcacag 1052
          ||||||||| |||||||| ||||||| |||||| ||||||||||||| |||||||||
Db   4049 ctggtctcctccctcaaaacctcgacttcacctccaagg------gagcag 4102

Qy   1053 tagcaacaacacaaaccacagcagctggctggagatcgattgggggagaaaagat 1112
          |||||| ||||| |||| ||||| |||| || ||| ||||||||| |||| |||||
Db   4103 acctcagtgcgataatccaaaagagtggctgcaagtggacttccaagaacatgaaagt 4162

Qy   1113 aacaggaattaggaccacagatctacacagtcgaactcaacttctattgttaagaggtt 1172
          ||||||| | ||| ||||| || | |||| |||| || ||| ||||||| || || |||
Db   4163 cacagagttactactcaggggtaaaatctctgcttaccagcatgtatgtcgaagagtt 4222

Qy   1173 tgtgatgaacttcaaaagaacatattctacctaaaggaaattgtgaataa 1232
          ||| ||| ||||| || ||||| | ||||| ||||| |||||||||| ||
Db   4223 cctcatctccagcacgttaaagatggacgggttgtcccacagacatgggacaggaagatagc 4276

Qy   1233 tagcaacaaccacaaaccacagaggatctacacagtcgaactcaact 1292
          ||| | || | | ||| ||||||| |||| |||| | |||| ||||||
Db   4277 caaagtaaaagttttcaggaaacctggaacaaaatggacaggaagatagc 4336

Qy   1293 cctcccatcctggccagatatgtgcgggttgtcccacagacaagaggatagc 1352
          |||||| |||||||| ||| || | | ||||| |||||||| ||||||||||
Db   4337 ccctacccgttactgactcgctacttccaccctggagaaagaaattcat 4396

INFORMATION FOR SEQ ID NO: ...

Query Match   4.5%;   Score 78.4;   DB 4;   Length 4999;
Best Local Similarity   54.1%;   Pred. No. 4 1e-10;
Matches 210;   Conservative   0;   Mismatches 166;   Indels 12;   Gaps 2;

Qy   993 ctggtctcctggccaagccgacttcaggaccaagcgccatcggcttcggcgcacag 1052
          ||||||||| |||||||| ||||||| |||||| ||||||||||||| |||||||||
Db   4396 ctggtctcctccctcaaaacctcgacttcacctccaagg------gagagtaatgcctggag 4449

Qy   1053 tagcaacaacacaaaccacagcagctggctggagatcgattgggggagaaaagat 1112
          |||||| ||||| |||| ||||| |||| || ||| ||||||||| |||| |||||
Db   4450 acctcagtgaataatccaaaagagtggctgcaagtggacttccaagaacatgaaagt 4509

Qy   1113 aacaggaattaggaccacagatctacacagtcgaactcaacttctattgttaagaggtt 1172
          ||||||| | ||| ||||| || | |||| |||| || ||| ||||||| || || |||
Db   4510 cacagagttactactcaggggtaaaatctctgcttaccagcatgtatgtcgaagagtt 4569

Qy   1173 tgtgatgaacttcaaaagaacatattctacctaaaggaaattgtgaataa 1232
          ||| ||| ||||| || ||||| | ||||| ||||| |||||||||| ||
Db   4570 cctcatctccagcacgttaaagatggacgggttgtcccacagacaagaggatagc 4623

Qy   1233 tgaagaaaggtgttcaggaaacctggaacaaaatggacaggaagatagc 1292
          || | || | | ||| ||||||| |||| |||| | |||| ||||||||
Db   4624 caaagtaaaagttttcaggaaacctggaacaaaatggacaggaagatagc 4683

Qy   1293 cctcccatcctggccagatatgtgcgggttgtcccacagacaagaggatagc 1352
          |||||| |||||||| ||| || | | ||||| |||||||| ||||||||||
Db   4684 cccaccgttactgactcgctacttccaccctggagaaagaaattcat 4743

Qy   1353 cttgaagtggagctcattggtcggcgag 1380
          |||| | ||||| ||||| | | ||| |
Db   4744 cctgagatggaggttctgggctgcgag 4771

APPLICANT: Colosi, Peter C.
TITLE OF INVENTION: ADENO-ASSOCIATED VECTORS FOR EXPRESSION OF FACTOR VIII
TITLE OF INVENTION: BY TARGET
TITLE OF INVENTION: CELLS
FILE REFERENCE: AVIGEN-03743
CURRENT APPLICATION NUMBER: US/09/364,862
CURRENT FILING DATE: 1999-07-30
EARLIER APPLICATION NUMBER: 60/125,974
EARLIER FILING DATE: 1999-03-24
EARLIER APPLICATION NUMBER: 60/104,994
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 14
LENGTH: 4999
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-364-862-14

Query Match   54.1%; Score 78.4; DB 4; Length 4999;
Best Local Similarity   54.1%; Pred. No. 4.1e-10;
Matches 210; Conservative   0; Mismatches 166; Indels 12; Gaps 2;

```
Qy   993 ctggtctcctggccaaggccggacttcaaggccaatcatggcttcgggcgcacag 1052
             |||||||     |||||||| |||||| | ||||                |||||
Db  4396 ctggtctcctcaaaagctcgacttcactccagg------gaggagtaatgctggg 4449

Qy  1053 tagcacacaacacaaacagacgagtggctgggagaatcgatttgggggagaaaagaaaat 1112
             |||| |||| || |||| ||  ||||||| |||| |  |||| |||||||| || ||
Db  4450 accttcaggtgaataatccacaaaaaagaagttggctgcaagtggacttccagagaacaatgaaagt 1112

Qy  1113 aacaggaattaggacacagagactcacacagtcgaactcaacttttatctctaaagtt 1172
             |||||  ||| || |  ||| ||||  ||||| |||||| |||  | |||||||||
Db  4510 cacaggagtaactactcaggagtaaatctctgctctaccacgtatgtgaaggagtt 4569

Qy  1173 tgtgtgaacttcaaaacacataattctaaagtggagagacctataaaggaattgtgaataa 1232
             || ||| |||||| |  ||| ||| ||  |||||  |||||  | |||| ||  |||||
Db  4570 cctcatccagcgtcaagatggccatcagtggactctcttt----ttcagaatgg 4623

Qy  1233 tgaagaaaagtgttcaggtgaactcaacttcggagacccagtgcaaaacaatttcat 1292
             || ||||| |||||||| |||||| |||| || || | | |  |||||| |||||||
Db  4570 cacaggagttactctgcaggagctaaaatctctgctaccagctgtatgtgaaggagtt 4800

Qy  1293 ccctccatcgctggccagatatgtggtgttgtcccccagacatggcaacaggatagc 1352
             ||||| || |||||| || | |||| | ||||| |||||| | |||  ||||| |||
Db  4624 caaggtgaaagtgttccagggaaatcaagacatcctcaacactgtcgtgaactcctaga 4683

Qy  1353 cttgaaggtggagagctcattggttgccag 1380
             |||| || ||| |  | |||| |||| |||
Db  4684 ccacccgttactgactcgaattcacccccaagagttgggtcgccccagatgtc 4743

Db  4744 cctgagatgtggaggttctgggctggcagg 4771
```

RESULT 10
US-08-882-083-1
Sequence 1, Application US/08882083
Patent No. 5669292
GENERAL INFORMATION:
APPLICANT: VOORBERG, Johannes J.
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,083
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 514
FILING DATE: 13-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: ISACSON, John P.
REGISTRATION NUMBER: 33,715
REFERENCE/DOCKET NUMBER: 30472/212
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5399
TELEFAX: (202)672-5500
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 35..5017
US-08-882-083-1

Query Match   54.1%; Score 78.4; DB 2; Length 5035;
Best Local Similarity   54.1%; Pred. No. 4.1e-10;
Matches 210; Conservative   0; Mismatches 166; Indels 12; Gaps 2;

```
Qy   993 ctggtctcctggccaaggccggacttcaaggccaatcatggcttcgggcgcacag 1052
             |||||||||||    |||||||  ||||| |||| | |||          |||||
Db  4627 ctggtctcctcaaaagctcgacttcactccagg------gaggagtaatgctgggg 4680

Qy  1053 tagcacacaacacaaacagacgagtggctgggagatcgatttgggggagaaaaagaaat 1112
             |||| |||| ||| || |||| |||||| |||||  | |||| ||||||| | ||||
Db  4681 accttcaggtgaataatccaaaaaagaagttggctgcaagtggacttccagagaacaatgaaagt 4740

Qy  1113 aacaggaattaggacacagagactcacacagtcgaactcaacttttatctctaaagtt 1172
             |||||| ||| || |  ||| ||||  ||||| ||||||| |||   ||||||||||
Db  4741 cacaggagtaactactcaggagtaaatctctgctctaccacgtatgtgaaggagtt 4800

Qy  1173 tgtgtgaacttcaaaacacataattctaaagtggagagacctataaaggaattgtgaataa 1232
             || ||| ||||| |  |||| ||| ||  |||| |  |||| |  ||||| ||  |||||
Db  4801 cctcatccagcgtcaagatggccatcagtggactctcttt----ttcagaatgg 4854

Qy  1233 tgaagaaaagtgttcaggtggtcaacttaactcacttcggagacccagtgcaaaacaatttcat 1292
             || ||||| |||||||| || | ||||| |||   |||| || ||| | |  ||||| |||||||
Db  4855 cacaggagttactctgcaggagctaaaatctctgctaccagctgtatgtgaaggagtt 4914

Qy  1293 ccctccatcggccagatatgtgcgggttgtccccagacactggcgaccaggaggatagc 1352
             ||||| ||| ||||| || | |||| | || ||||| | ||||  |||||  |||| |||
Db  4915 cccaccgttactgactcgctacctctgaattcaccccccagagttgggcacaccagaatgg 4974

Qy  1353 cttgaaggtggagagctcattggttgccag 1380
             |||| || ||| || | |||| |||| |||
Db  4975 cctgaggatgtggaggttctgggctgcgcag 5002
```

RESULT 11
US-08-558-107-1
Sequence 1, Application US/08558107
Patent No. 5910481
GENERAL INFORMATION:
APPLICANT: VOORBERG, Johannes J.
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500

;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20007-5109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/558,107
;   FILING DATE: 13-NOV-1995
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: ISACSON, John P.
;   REGISTRATION NUMBER: 33,715
;   REFERENCE/DOCKET NUMBER: 30472/212
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202)672-5300
;   TELEFAX: (202)672-5399
;   TELEX: 904136
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5035 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 35..5017

US-08-558-107-1

Query Match   4.5%;   Score 78.4;   DB 2;   Length 5035;
Best Local Similarity   54.1%;   Pred. No. 4.1e-10;
Matches 210;   Conservative   0;   Mismatches 166;   Indels 12;   Gaps 2;

```
Qy    993 ctgtgtctcctggccaagcccgacttcaggaccaaggcccatcatggccttgggcgacag 1052
          ||||||||||   |||
Db   4627 CTGGTCTCCTTCAAAAGCTCGACTTCACCTCCAAGG------GAGGAGTAATGCCTCGGAG 4680

Qy   1053 tagcaacaacacacaaaccagagagttggctgggagtcatttgggaggaaaagagaaat 1112
Db   4681 ACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTATGTGAGGAGT 4740

Qy   1113 aacaggaattaggaccacaggatctacacagtcgaacttcaacttttatgttaagagtt 1172
Db   4741 CACAGGAGTAACTACTCAGGGAGTAAAATCTCGCTTACCACGATGTATGTCAGAGGAGTT 4800

Qy   1173 tgtgatgaacttcaaaaacaatattcctaagtggaagacctataaaggaattgtgaataa 1232
Db   4801 CCTCACTCCACAGAGTCAAGATGCCCATCAGTGGACTCTCTTTT-----TTCAGAATGG 4854

Qy   1233 tgaagaaaaggtgttcaggtaacttaactcaacttcgggaccacgtgcaaaacaattcat 1292
Db   4855 CAAAGTAAAAGTTTTTCAGGGAAATCAACAGACTCCTTCACACCCTGGGTGCCCAGAGT 4914

Qy   1293 ccctcccatcgtggccagatatgtgcgggttgtccccagacaatggcaggagatagc 1352
Db   4915 CCACCGTTACTGACTCGCTACCTTGGAATTCACCCCCAGAGCTGGGTGGCCAGAGT 4974

Qy   1353 cttgaagtggaagctcattggttgccag 1380
Db   4975 CCTCAGGATGGAGGTTCTGGGCTGCGAG 5002
```

RESULT 12
US-09-243-539-1
Sequence 1, Application US/09243539
Patent No. 6130203
GENERAL INFORMATION:
APPLICANT: VOORBERG, Johannes J.
TITLE OF INVENTION: HYBRID PROTEINS WITH MODIFIED ACTIVITY

;   NUMBER OF SEQUENCES: 17
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 3000 K Street, N.W., Suite 500
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20007-5109
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/243,539
;   FILING DATE:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/558,107
;   FILING DATE: 13-NOV-1995
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: ISACSON, John P.
;   REGISTRATION NUMBER: 33,715
;   REFERENCE/DOCKET NUMBER: 30472/212
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202)672-5300
;   TELEFAX: (202)672-5399
;   TELEX: 904136
;   INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5035 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 35..5017

US-09-243-539-1

Query Match   4.58%;   Score 78.4;   DB 3;   Length 5035;
Best Local Similarity   54.1%;   Pred. No. 4.1e-10;
Matches 210;   Conservative   0;   Mismatches 166;   Indels 12;   Gaps 2;

```
Qy    993 ctgtgtctcctggccaagcccgacttcaggaccaaggcccatcatggccttgggcgacag 1052
          ||||||||||   |||
Db   4627 CTGGTCTCCTTCAAAAGCTCGACTTCACCTCCAAGG------GAGGAGTAATGCCTCGGAG 4680

Qy   1053 tagcaacaacacacaaaccagagagttggctgggagtcatttgggaggaaaagagaaat 1112
Db   4681 ACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTATGTGAGGAGT 4740

Qy   1113 aacaggaattaggaccacaggatctacacagtcgaacttcaacttttatgttaagagtt 1172
Db   4741 CACAGGAGTAACTACTCAGGGAGTAAAATCTCGCTTACCACGATGTATGTCAGAGGAGTT 4800

Qy   1173 tgtgatgaacttcaaaaacaatattcctaagtggaagacctataaaggaattgtgaataa 1232
Db   4801 CCTCACTCCACAGAGTCAAGATGCCCATCAGTGGACTCTCTTTT-----TTCAGAATGG 4854

Qy   1233 tgaagaaaaggtgttcaggtaacttaactcaacttcgggaccacgtgcaaaacaattcat 1292
Db   4855 CAAAGTAAAAGTTTTTCAGGGAAATCAACAGACTCCTTCACACCCTGGGTGCCCAGAGT 4914

Qy   1293 ccctcccatcgtggccagatatgtgcgggttgtccccagacaatggcaggagatagc 1352
Db   4915 CCACCGTTACTGACTCGCTACCTTGGAATTCACCCCCAGAGCTGGGTGGCCAGAGT 4974

Qy   1353 cttgaagtggaagctcattggttgccag 1380
Db   4975 CCTCAGGATGGAGGTTCTGGGCTGCGAG 5002
```

RESULT 13
US-08-276-594A-1
Sequence 1, Application US/08276594A
Patent No. 5693499
GENERAL INFORMATION:
  APPLICANT: YONEMURA, Hiroshi
  APPLICANT: TAJIMA, Yoshitaka
  APPLICANT: SUGAWARA, Keishin
  APPLICANT: MASUDA, Kenichi
  TITLE OF INVENTION: PROCESS FOR PREPARING HUMAN COAGULATION
  TITLE OF INVENTION: FACTOR VIII PROTEIN COMPLEX
  NUMBER OF SEQUENCES: 11
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Foley & Lardner
    STREET: 3000 K Street, N.W., Suite 500
    CITY: Washington
    STATE: D.C.
    COUNTRY: USA
    ZIP: 20007-5109
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/276,594A
    FILING DATE: 18-JUL-1994
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/950,191
    FILING DATE: 24-SEP-1992
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 243262/1991
    FILING DATE: 24-SEP-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: WEGNER, Harold C.
    REGISTRATION NUMBER: 25,258
    REFERENCE/DOCKET NUMBER: 74129/195/AOPA
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (202)672-5300
    TELEFAX: (202)672-5399
    TELEX: 904136
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 6999 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..6996

US-08-276-594A-1

Query Match   54.5%;   Score 78.4;   DB 1;   Length 6999;
Best Local Similarity   54.1%;
Matches 210;   Conservative   0;   Mismatches 166;   Indels 12;   Gaps 2;

```
Qy    1053 ctggtctcctggccaagcccgacttcaggagcccc................
           | ||||||||||||||||||||||||||||||||
Db    6660 ACCTCAGGTGAATAAATCCAAAGAGAGTAGGGTCTGCAACTGACTTCGACAAGAAGAAAGT 6719
Qy    1113 aacagagaattaggaccaaggatctacacagtcgaacttcaactttatgttaagagtt 1172
           |||| ||||||  | ||||||||||||||||||||||
Db    6720 CACAGGACTAAACTACTCAGGGAGTAAAATCTCTGCTTACCAGCNNTGTATNGTGAAGCNAGTT 6779
Qy    1173 tgtgatgaacttcaaaaacaataattctaagttggaagacctataaaggagattgtgaataa 1232
           ||||||||||| |||| | ||||| |||||||||
Db    6780 CCCTCATCCTCCAGCAGTCAGTCAAGATGGCCATCAGTGGCCACTCTCTTTT---TTCAGAATGG 6833
```

RESULT 14
US-08-121-202-1
Sequence 1, Application US/08121202
Patent No. 5563045
GENERAL INFORMATION:
  APPLICANT: Pittman, Debra
  APPLICANT: Rehemtulla, Alnawaz
  APPLICANT: Wozney, John M.
  APPLICANT: Kaufman, Randal J.
  TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS
  NUMBER OF SEQUENCES: 27
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
    STREET: 87 CambridgePark Drive
    CITY: Cambridge
    STATE: MA
    COUNTRY: USA
    ZIP: 02140
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/121.202
    FILING DATE: 14-SEP-1993
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Meinert, M. C.
    REGISTRATION NUMBER: 31,544
    REFERENCE/DOCKET NUMBER: GI 5195A
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (617)876-1210 X8574
    TELEFAX: (617) 876-5851
  INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 7056 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: cDNA
    HYPOTHETICAL: NO
    ANTI-SENSE: NO
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..7053

US-08-121-202-1

Query Match   54.5%;   Score 78.4;   DB 1;   Length 7056;
Best Local Similarity   54.1%;
Matches 210;   Conservative   0;   Mismatches 166;   Indels 12;   Gaps 2;

```
Qy    1053 ctggtctcctggccaagcccgacttcaggagcccc................
           | ||||||||||||||||||||||||||||||||
Db    6663 CTGGTGTCCTTCAAAAGCTCGACTTCACCTCCACAAGG----GAGGAGTAATGCCTGGAG 6716
Qy    1053 tagcaacaacacaaccacgagagaagtggctggaagatcgattgggggaggaaaaagagaaat 1112
```

```
RESULT 15
5171844-1
;Patent No. 5171844
;APPLICANT: VAN OOYEN, ALBERT J.J.;PANNEKOEK, HANS;VERBEET,
;MARTINUS P.;VAN LEEN, ROBERT W.
;TITLE OF INVENTION: PROTEINS WITH FACTOR VIII ACTIVITY
;PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS
;AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM
;NUMBER OF SEQUENCES: 12
;CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/07/205,226
;  FILING DATE: 10-JUN-1988
;SEQ ID NO:1:
;  LENGTH: 8241

Query Match        4.5%;   Score 78.4;  DB 6;   Length 8241;
Best Local Similarity   54.1%;   Pred. No. 5.1e-10;
Matches  210;  Conservative    0;  Mismatches  0;  Indels  12;  Gaps  2;

Qy     993 ctggtctcctggccaaggcccgacttcaggagaccaaggcccatcatggggtccgggcgacag 1052
           ||||||||||  || ||||||||||| |||| || |||||  ||| ||||||
Db    6669 ctggtctcctcaaaaggctcgacttcaacctcaagg------gaggagtaatgcctggag 6722

Qy    1053 tagcaacacacacaacaccacgagagtgtgctgaagtcgattgggggagaaaaagagaaat 1112
           | |||| ||| ||  ||||| ||| ||| ||| |||| || |||| |||||| ||
Db    6723 acctcaggtcgaataatccaaaaggtggctgcaagtggacttccagagacacgaaagt 6782

Qy    1113 aacaggaattaggaccacaggatctacacgagtcgaacttcaactttttatgttaagagtt 1172
           |||| |||| ||||||||| |||||| || || |||||||||||| |  ||| |||||||
Db    6783 cacaggagtaactactcaggagtaaaatctcgcttaccagcatgtatgtgaaggagtt 6842

Qy    1173 tgtgtgatgaacttcaaaaacaataattctaagtgaagaccctataaaggaattgtgaataa 1232
           || |||||| ||||||| ||| || ||||| |||| ||||||| |||||||| ||
Db    6843 cctcatctccagccagtcaagatggccatcaggtggactctctttt------ttcagaatgg 6896

Qy    1233 tgaagaaaagtgtttcaggggtaactctaacttcgggacctataaaggcaaaacaattcat 1292
           ||||||||| || ||||||||||| ||| ||||| ||||| |||| ||| || ||||| ||
Db    6897 caaagtaaagtgttttcaggggaatcaagactccttcacacctgtgaactctctaga 6956

Qy    1293 ccctcccatcgtggccagatatgtgcgggttgtccccacgacagaaggatagc 1352
           ||| ||| ||| ||| |||  ||||| ||||| ||||| ||| || ||||||| ||
Db    6957 ccacacgttactgactcgctacccttcgaattcaccccccagaggttgggtgcaccagattgc 7016

Qy    1353 cttgaaggtcgagctcattggttgcag 1380
           |||||  | | ||||  ||| |||||
Db    7017 cctgaggatcggaggttctggggctgcgcgag 7044
```

```
Db    6717 ACCTCAGGTCAGGTGAATAATCCAAAAAGAGTGGCCTGCCAAGTGGACTTCACTTTCCAGAGACAATCGAAAGT
              |||||||||  || ||||||||||| |||| || |||||  ||| ||||||
Qy    1113 aacaggaattaggaccacaggatctacacagtgcccaagtcgacacttttatgttaggagtt 1172
Db    6777 CACAGGAGTAACTACTCAGGAGTAAAATCTCGTTACCAGCATGTATGTGAAGGAGTT 6836
Qy    1173 tgtgtgatgaacttcaaaaacaataatctaagtctcggagagaccatataaaggaattcgtgaataa 1232
              |||||| ||| ||||| || |||| | || ||| ||||||  || ||| ||||||  ||| |||
Db    6837 CCTCATCTCCAGCCAGTCAAGATGGCCATCACGTGGACTCTCTTTT------TTCAGAATGG 6890
Qy    1233 tgaagaaaagtgtttcaggggtaactctaacttcgggacccagtcgcaaaacaattcat 1292
              |||| ||||| || ||||||||| |  ||||  ||||| ||| ||| || ||||| ||
Db    6891 CAAAGTAAAGTGTTTTCAGGGAAATCAAGACTCCTTCACACCTGTGGTCGAACTCTCTAGA 6950
Qy    1293 ccctcccatcgtggccagatatgtgcgggttgtccccacgacagaaggatagc 1352
              |||  || | ||| |||  ||| | |||| || |||| ||| ||  |||||||  |
Db    6951 CCCACCGTTACTGACTCGCTACCCTTCGAATTCACCCCCCAGAGAGTTGGGTGCACCAGATTGC 7010
Qy    1353 cttgaaggtcgagctcattggttgcag 1380
              |||| | |  |||| ||| | |||||
Db    7011 CCTGAGGATCGGAGGTTCTGGGGCTGCCAG 7038
```

OM protein - protein search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: July 9, 2002, 01:07:13 ; Search time 80.25 seconds

Title: US-09-691-344A-4
Perfect score: 3052
Sequence: 1 MGFGAGQRLRPVPAPRSSAE........YDNEKEMTQKLDLITSDMAG 586

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A.Geneseq_032802:*
                 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

811.081 Million cell updates/sec
(without alignments)

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	3052	100.0	586	22	AAU00629	Novel human protei
2	2818	92.3	539	22	AAU00630	Novel human protei
3	2812	92.1	715	22	AAU00670	Human semaphorin
4	2557	83.8	487	22	AAU00628	Human TANGO 229 po
5	1301.5	42.6	503	21	AAB19126	Novel human protei
6	903.5	29.6	583	21	AAG75450	Polypeptide isolat
7	786.5	25.8	669	21	AAY70539	Human colon cancer
8	353.5	11.6	889	21	AAB24216	Human Factor 8 Hom
9	351.	11.5	925	20	AAW96308	Soluble neuropili
10	349	11.4	439	22	AAB64627	Neuropilin-2, Rat
11	349	11.4	555	20	AAT06520	Human soluble neur

ALIGNMENTS

RESULT 1
ID   AAU00629 standard; protein; 586 AA.
XX
AC   AAU00629;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Novel human protein (NHP) sequence #2.
XX
KW   Novel human protein; NHP; CUB domain; extracellular domain;
KW   obesity; high blood pressure; connective tissue disorder; infertility;
KW   NHP-mediated pathway.
XX
OS   Homo sapiens.
XX
PN   WO200129219-A1.
XX
PD   26-APR-2001.
XX
PF   08-OCT-2000; 2000WO-US28798.
XX
PR   19-OCT-1999; 99US-0160285.
PR   18-FEB-2000; 2000US-0183583.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands A?;
XX
DR   WPI; 2001-290917/30.
DR   N-PSDB; AAS00614.
XX
PT   Novel nucleic acid encoding human CUB-domain containing protein, useful
PT   for drug screening, diagnosis and treatment of physiological disorders

12	349	11.4	20	AAW96254	Mouse semaphorin r
13	349	11.4	20	AAW96255	Human semaphorin r
14	349	11.4	20	AAY06318	Human neuropilin-2
15	349	11.4	20	AAY14563	Human VEGF(165),R/N
16	349	11.4	20	AAY23248	Human NP-2 protein
17	345.5	11.3	22	AAB64570	Human secreted pro
18	343.5	11.3	20	AAW96256	Mouse semaphorin r
19	343.5	11.3	20	AAW96257	Human semaphorin r
20	343.5	11.3	20	AAW96249	Rat semaphorin rec
20	343.5	11.3	20	AAW96250	Human semaphorin r
21	343.5	11.3	20	AAW96244	Mouse semaphorin r
22	343.5	11.3	20	AAW96251	Human semaphorin r
23	343.5	11.3	20	AAW96252	Mouse semaphorin r
24	343.5	11.3	20	AAW96253	Human semaphorin r
25	341	11.2	22	AAM25770	Human protein sequ
26	439	22		AAB64628	Human secreted pro
27	538	22		AAU02950	Angiotensin conver
28	600	22		AAU02948	Angiotensin conver
29	644	22		AAY06319	Angiotensin conver
30	840	22		AAU02949	Human soluble neur
31	856	20		AAU02949	Angiotensin conver
32	892	21		AAY06317	Angiotensin conver
33	923	20		AAB62478	Human soluble neur
34	923	20		AAY23247	Angiotensin conver
35	923	20		AAW96246	Human semaphorin r
36	922	20		AAB24213	Human neuropilin-1
37	924	20		AAW96309	Human neuropilin-1
38	923	20		AAW96248	Human VEGF165R/NP-
39	923	20		AAW96309	Human neuropilin-1
40	923	21		AAB24213	Human semaphorin r
41	921	20		AAW96247	Mouse soluble neur
42	921	21		AAY49954	Rat semaphorin rec
43	2183	21		AAB03533	Neuropilin, Rattu
44	2224	21		AAW04254	Murine factor V SE
44	2224	17		AAW04254	Murine factor V, H
45	159	22		AAB71907	Human FV C2 domain

Left column

```
PT   or diseases -
XX
XX
PS   Claim 2; Page 27-28; 33pp; English.
XX
XX
CC   The sequence represents a novel human protein (NHP) containing a CUB
CC   domain (an extracellular domain). CUB proteins have been associated with
CC   regulating development, modulating cellular processes and preventing
CC   infectious disease. NHP nucleotide sequences are useful for gene therapy
CC   of physiological disorders or diseases. NHP oligonucleotides are useful
CC   as hybridisation probes for screening libraries and assessing gene
CC   patterns. NHP nucleotide sequences are useful for detecting mutant or
CC   inappropriately expressed NHPs (for example, those proteins associated
CC   with obesity, high blood pressure, connective tissue disorders and
CC   infertility) for the diagnosis of a disease. The polynucleotides may also
CC   be used in screening for drugs effective in the treatment of symptomatic
CC   or phenotypic manifestations of perturbing the normal function of NHP in
CC   the body. Nucleotide constructs encoding NHP products are used to
CC   genetically engineer host cells to express such products in vivo. These
CC   host cells allow for the identification of compounds that bind to NHP
CC   receptors or trigger NHP-mediated pathways.
XX
SQ   Sequence      586 AA;
XX
```

```
Query Match            100.0%;  Score 3052;  DB 22;  Length 586;
Best Local Similarity  100.0%;  Pred. No. 5.7e-265;
Matches  586;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 MGFGAQORLRLRPVPAPRSSAERANRPGCQLRLGIRRGEAELAKLAPSGVMWPGCARGGALAR 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 mgfgaqqrlrlrpvpapraaeeaaarpgcqlrlgirrgeaelaklapsgvmwpgcarggalar 60

Qy    61 AAGRGLLALLLAVSAPLRLQAEELGDGCGHLVTYQDSGTMTMSKNYPGTYPNHTVCEKTIT 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 aagrgllalllavsaplrlqaeelgdgcghlvtyqdsgtmtmsknypgtypnhtvcektit 120

Qy   121 VPKGKRLILRLGDLLDIESQTCASDYLLFTSSSDQYGCYCGSMYVPKELLLNTSEVTVRFE 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 vpkgkrlilrlgdlldiesqtcasdyllftsssdqygcycgsmyvpkelllntsevtvrfe 180

Qy   181 SGSHLSGRGFLLTYASSDHPDLITCLERASHYLKTEYSKFCPAGCRDVAGDISGNNVPGY 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 sgshlsgrgflltyassdhpdlitclerashylkteyskfcpagcrdvagdisgnnvpgy 240

Qy   241 RDTSLLCKAAIHAGIIADELGGQISVLQRKGISRYEGILLANGVLSRDGSLSDKRFLFTSN 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 rdtsllckaaihagiiadelggqisvlqrkgisryegillangvlsrdgslsdkrflftsn 300

Qy   301 GCSRSLSFEPDGIRASSSWQSVWESGDDVHNSPGQQARLQDQGPSWASGGDSSNNIKPREW 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 gcsrslsfepdgirassswqsvwesgddvhnspgqqarlqdqgpswasggdssnnikprew 360

Qy   361 LEIDLGEKKKWIIGIRTTGSTQSNFNFYVKSFVWNFRKNNNSKWKTTYKGIVNNEEKVFQGNS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 leidlgekkkwiigirttgstqsnfnfyvksfvwnfrknnnskwktyk givnneekvfqgns 420

Qy   421 NFRDPVQNNEFIPFIVARYRVVPQTWDHRIALKVELIGCCITGGNDSLVWRKTSQSTSVS 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 nfrdpvqnnfippivaryrvvvpqtwdhqrialkveligcqitqgndslvwrktsqstsvs 480

Qy   481 TKKEDETITIRPIPSEETSTGINITTVAIPLVLLVVLVPAGMGIFAAFRKKKKKGSPYCGSA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 tkkedetitrpipseetstginittvaipivlvlvlvfagmgifaafrkkkkkgspygsa 540

Qy   541 EAQKTDCWKQIKYPFARHQSAEFTISYDNEKEMTQKDLITSDMAG 586
           ||||||||||||||||||||||||||||||||||||||||||||
Db   541 eaqktdcwkqikypfarhqsaeftisydnekemtqkdlitsdmag 586
```

Right column

```
AC   AAU00630;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Novel human protein (NHP) sequence #3.
XX
KW   Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW   obesity; high blood pressure; connective tissue disorder; infertility;
KW   NHP-mediated pathway.
XX
OS   Homo sapiens.
XX
CC   WO200129219-A1.
XX
PN   26-APR-2001.
XX
PD   08-OCT-2000; 2000WO-US28798.
XX
PF   19-OCT-1999; 99US-0160285.
PF   18-FEB-2000; 2000US-0183583.
XX
PT   N-PSDB; AAS00615.
DR   WPI; 2001-290917/30.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Donoho G,  Turner CA,  Nehls M,  Friedrich G,  Zambrowicz B,  Sands AT;
XX
XX
XX
XX
PS   Claim 5; Page 29-30; 33pp; English.
XX
XX
CC   The sequence represents a novel human protein (NHP) containing a CUB
CC   domain (an extracellular domain). CUB proteins have been associated with
CC   regulating development, modulating cellular processes and preventing
CC   infectious disease. NHP nucleotide sequences are useful for gene therapy
CC   of physiological disorders or diseases. NHP oligonucleotides are useful
CC   as hybridisation probes for screening libraries and assessing gene
CC   patterns. NHP nucleotide sequences are useful for detecting mutant or
CC   inappropriately expressed NHPs (for example, those proteins associated
CC   with obesity, high blood pressure, connective tissue disorders and
CC   infertility) for the diagnosis of a disease. The polynucleotides may also
CC   be used in screening for drugs effective in the treatment of symptomatic
CC   or phenotypic manifestations of perturbing the normal function of NHP in
CC   the body. Nucleotide constructs encoding NHP products are used to
CC   genetically engineer host cells to express such products in vivo. These
CC   host cells allow for the identification of compounds that bind to NHP
CC   receptors or trigger NHP-mediated pathways.
XX
SQ   Sequence      539 AA;
XX
```

```
AC   AAU00630;
XX
ID   AAU00630 standard; Protein; 539 AA.
XX
DB   Novel nucleic acid encoding human CUB-domain containing protein, useful
Qy   for drug screening, diagnosis and treatment of physiological disorders
Db   or diseases -
```

```
Query Match            92.3%;  Score 2818;  DB 22;  Length 539;
Best Local Similarity  100.0%;  Pred. No. 5.e-244;
Matches  539;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    48 MWPGCARGGALARAAGRGLLALLLAVSAPLRLQAEELGDGCGHLVTYQDSGTMTMSKNYPG 107
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 mwpgcarggalaraagrgllalllavsaplrlqaeelgdgcghlvtyqdsgtmtmsknypg 60

Qy   108 TYPNHTVCEKTITVPKGKRLILRLGDLLDIESQTCASDYLLFTSSSDQYGCYCGSMYVPKE 167
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 typnhtvcektitvpkgkrlilrlgdlldiesqtcasdyllftsssdqygcycgsmyvpke 120

Qy   168 LLLNTSEVTVRFESGSHLSGRGFLLTYASSDHPDLITCLERASHYLKTEYSKFCPAGCRD 227
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 lllntsevtvrfesgshlsgrgflltyassdhpdlitclerashylkteyskfcpagcrd 180

Qy   228 VAGDISGNNVPGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRYEGILLANGVLSRD 287
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 vagdisgnnvpgyrdtsllckaaihagiiadelggqisvlqrkgisryegillangvlsrd 240
```

[Handwritten annotation: SEQ ID NO: 4 Database: A_Geneseq-032802 AC NO: AAU00670]

RESULT 3

```
AA006670
ID   AAU00670 standard; Protein; 715 AA.
XX
AC   AAU00670;
XX
DT   07-SEP-2001 (first entry)
XX
DE   Human TANGO 229 polypeptide.
XX
KW   Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
KW   skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW   bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW   anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease;
KW   malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW   attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW   hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW   muscular dystrophy; immuno-competence; vertebrate; blood; serum.
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FH
FT   Peptide        1..34
FT                  /note= "signal peptide"
FT   Domain         35..455
FT                  /note= "Extracellular domain"
FT   Protein        35..715
FT                  /note= "Mature human TANGO 229"
FT   Domain         456..480
FT                  /note= "Transmembrane domain"
FT   Domain         481..715
FT                  /note= "Cytoplasmic domain"
XX
PN   WO200129088-A1.
XX
PD   26-APR-2001.
XX
PF   23-JUN-2000; 2000WO-US17386.
XX
PR   19-OCT-1999; 99US-0420707.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Mackay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
XX
DR   WPI; 2001-308477/32.
DR   N-PSDB; AAU00660.
XX
PT   New isolated nucleic acid molecule for diagnosis, prevention, and
PT   therapy of human and other animal disorder, or as modulating agent for
PT   regulating cellular processes
```

```
Qy  288  GSLSDKRFLFTSNCCSRSLSFEPDQQIRASSMQSVNESGDQVHWSPGQARLQDQGPSWA  347
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  gslsdkriflftsngcsrslsfepdgqqirasswqsvnesgdqvhwspgqarlqdqgpswa  300

Qy  348  SGDSSNHKPREWLEIDLGEKKKTIGIRTTGSTQSNRNEFYVKSFYWNNEKNNNSKWKTYKG  407
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  sgdssnnhkprewleidlgekkktigirttgstqsnrnfyyksfvmnfknnnskwktykg  360

Qy  408  IVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELIGCCITQGNDS  467
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ivnneekvfqgnsnfrdpvqnnfippivaryvrvvpqtwhqrialkveligcqitqgnds  420

Qy  468  LVWRKTSQSTSVSTKKEDEITTRPIPSEETSTGNITITVAIPLVILVVLVFAGMGIFAAF  527
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  lvwrktsqstsvstkkedeittrpipseetstginitvaipllvilvvlvfagmgifaaf  480

Qy  528  RKKKKKGSPYGSAEAQKTDCWKQIKYPFARHQSAEFTISTDNEKEMTQKLDLITSDMAG  586
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  rkkkkkgspygsaeaqktdcwkqikypfarhqsaeftisydnekemtqkldlitsdmag  539
```

```
XX
CC   Claim 8; Fig 1; 263pp; English.
XX
CC   The sequence represents human TANGO 229 polypeptide. This protein and
CC   similar others exhibit the ability to affect growth, proliferation,
CC   survival, differentiation, activity, morphology, or movement/migration
CC   of, e.g. T cells and cells of the heart, liver, pancreas, placenta,
CC   brain, lung, skeletal muscle, kidney, spleen, lymph node, peripheral
CC   blood leukocyte, bone marrow or thymus tissue. They can be used as
CC   modulating agents for regulating cellular processes, thus, the proteins
CC   and their associated nucleic acids can be used to prognosticate, prevent,
CC   diagnose or treat disorders associated with physiological processes.
CC   These disorders include abnormal blood coagulation, asthma, anaphylaxis,
CC   hepatitis, multiple sclerosis, cancer, coronary artery disease, malaria,
CC   atopic dermatitis, amyotrophic lateral sclerosis, meningitis, attention
CC   deficit disorder, Crohn's disease, gastroenteritis, goitre,
CC   hypoglycaemia, diabetes mellitus, endometriosis, pulmonary embolism and
CC   muscular dystrophy; immuno-competence; vertebrate; blood; serum.
CC   Antibodies to disorders such as these can be made by
CC   providing a polypeptide of the invention to an immuno-competent
CC   vertebrate and harvesting blood or serum from the vertebrate.
XX
SQ   Sequence 715 AA;
```

```
Query Match            92.1%;  Score 2812;  DB 22;  Length 715;
Best Local Similarity  100.0%;  Pred. No. 2.8e-243;
Matches  538;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy   48  MVPGARGGALARAAGRGLLALLLAVSAPLRLQAEELGDGCGHLVTYQDSGTMTSKNYPG  107
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  mvpgargggalaraegrglalllavsaplrlqaeelgdgcghlvtygdsgtmtsknypp   60

Qy  108  TYPNHTVCEKTITVPKRKIRLILRLGDLDITSQTCASDYLLFFSSSDYGPYCGSMTVPKE  167
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  typnhtvcektitvpkgkrlilliqdldlesqtcaadyllftsssdqypycgsmtvpke  120

Qy  168  LLLNTSEVYTVRFESSSHSRGFLLYASSDHPDLITCLERASHYLKTEYSKFCPACCRD  227
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  lllntsevtvrfsesghsrgfllfyyassdhpdlitclerashylkteyskfcpacerd  180

Qy  228  VAGDISGNMVDGYRDTSLLCKAAIHAGIIADEIGCCOISVLQRKGISRYEGILLANGVLSRD  287
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  vagdisgnmvdgyrdtsllckaaihagiiadelggqisvlqrkgisryegillangvlsrd  240

Qy  288  GSLSDKRFLFTSNCCSRSLSFEPDQQIRASSMQSVNESGDQVHWSPGQARLQDQGPSWA  347
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  gslsdkriflftsngcsrslsfepdgqqirasswqsvnesgdqvhwspgqarlqdqgpswa  300

Qy  348  SGDSSNHKPREWLEIDLGEKKKTIGIRTTGSTQSNRNEFYVKSFYWNNEKNNNSKWKTYKG  407
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  sgdssnnhkprewleidlgekkktigirttgstqsnrnfyyksfvmnfknnnskwktykg  360

Qy  408  IVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIALKVELIGCCITQGNDS  467
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ivnneekvfqgnsnfrdpvqnnfippivaryvrvvpqtwhqrialkveligcqitqgnds  420

Qy  468  LVWRKTSQSTSVSTKKEDEITTRPIPSEETSTGNITITVAIPLVILVVLVFAGMGIFAAF  527
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  lvwrktsqstsvstkkedeittrpipseetstginitvaipllvilvvlvfagmgifaaf  480

Qy  528  RKKKKKGSPYGSAEAQKTDCWKQIKYPFARHQSAEFTISTDNEKEMTQKLDLITSDMA  585
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  rkkkkkgspygsaeaqktdcwkqikypfarhqsaeftisydnekemtqkldlitsdma  538
```

RESULT 4

```
AAU00628
ID   AAU00628 standard; Protein; 487 AA.
XX
AC   AAU00628;
XX
DT   29-AUG-2001 (first entry)
XX
```

DE   Novel human protein (NHP) sequence #1.
XX
XX
KW   Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW   obesity; high blood pressure; connective tissue disorder; infertility;
KW   NHP-mediated pathway.
XX
OS   Homo sapiens.
XX
XX
PN   WO200129219-A1.
XX
PD   26-APR-2001.
XX
PF   08-OCT-2000. 2000WO-US287798.
DR   N-PSDB; AAS00613.
XX
PI   2001-299917/30.
PI   WPI;
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PA
PI   Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
PR
PR   19-OCT-1999;   99US-0160285.
PR   18-FEB-2000;   2000US-0183583.
XX
PT   Novel nucleic acid encoding human CUB-domain containing protein, useful
PT   for drug screening, diagnosis and treatment of physiological disorders
PT   or diseases
XX
PS   Claim 4; Page 26; 33pp; English.
CC
CC   The sequence represents a novel human protein (NHP) containing a CUB
CC   domain (an extracellular domain). CUB proteins have been associated with
CC   regulating development, modulating cellular processes and preventing
CC   infectious disease. NHP nucleotide sequences are useful for gene therapy
CC   of physiological disorders or diseases. NHP oligonucleotides are useful
CC   as hybridisation probes for screening libraries and assessing gene
CC   patterns. NHP nucleotide sequences are useful for detecting mutant or
CC   inappropriately expressed NHPs (for example, those proteins associated
CC   with obesity, high blood pressure, connective tissue disorders and
CC   infertility) for the diagnosis of a disease. The polynucleotides may also
CC   be used in screening for drugs effective in the treatment of symptomatic
CC   or phenotypic manifestations of perturbing the normal function of NHP in
CC   the body. Nucleotide constructs encoding NHP products are used to
CC   genetically engineer host cells to express such products in vivo. These
CC   host cells allow for the identification of compounds that bind to NHP
CC   receptors or trigger NHP-mediated pathways.
XX
XX
SQ   Sequence   487 AA;

Query Match   83.8%;   Score 2557;   DB 22;   Length 487;
Best Local Similarity   100.0%;   Pred. No. 1.2e-220;
Matches 487;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   100 MTSKNYPGTYPNEFVCEKTTIVPKGKRLIRLGLGDIDIESQTCASDYILFTSSDQYGPYC 159
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 mtsknypgtypnhfvcektitvpkgkrliritglgdidiesqtcasdyilftsssdqyypyc   60

Qy   160 GSMTVFKELLNTSEVVTVRFESGSHISGRGFLTYASSDHPDLITCLERASHYLKTEYSK 219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 gsmtvpkellntsevtvrfesgshisgrgfltyassdhpdlitclerashylkteysk 120

Qy   220 FCPAGCRDVAGDIGSGNMVDGYRDTSLLCKAIHAGIIADELGGQISVLQRKGISRYEGTL 279
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 fcpagcdvagdisgnmvdgyrdtsllckaihagiiadelggqisvlqrkgisryegtl 180

Qy   280 ANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGQTRASSKWQSVNESCDQVHHSPGQARL 339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 angvlsrdgslsdkrflftsngcsrslsfepdgqtrasskwqsvnesgdqvhhspgqarl 240

Qy   340 QDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTGSTQOSNFENFYVKKSFWMNRKNNN 399
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 qdqgpswasgdssnnhkprewleidlgekkkitgirttgstqsnfnfyvksfwmnfknn 300

SEQ ID NO; 4
Ac. NO.: AAB19126, Database; A.Genesq-0318_02

RESULT 5
ID   AAB19126 standard; protein; 503 AA.
XX
AC   AAB19126;
XX
DT   19-FEB-2001 (first entry)
XX
DE   Polypeptide isolated from lymph node stromal cells of fsn -/- mice.
XX
KW   Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW   immune system disorder; cancer; viral disorder; HIV infection;
KW   blood vessel growth; tumour necrosis factor disorder; arthritis;
KW   Inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW   cardiac failure.
XX
OS   Mus sp.
XX
XX
PN   WO200058463-A1.
XX
PD   05-OCT-2000.
XX
PF   18-FEB-2000. 2000WO-NZ00015.
XX
PR   25-MAR-1999;   99US-0276268.
PR   26-AUG-1999;   99US-0383586.
XX
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
PI   Murison JG;
XX
PI   WPI; 2000-664924/64.
DR   N-PSDB; AAS06736.
XX
PT   Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
PT   useful for modulating growth of blood cells, for treating inflammatory
PT   and tumour necrosis factor-mediated disorders, cancer and viral
PT   disorders
XX
PS   Claim 1; Page 68-69; 75pp; English.
CC
CC   The present sequence represents a polypeptide sequence which is
CC   isolated from lymph node stromal cells of fsn -/- mice. The
CC   polynucleotides and their polypeptides are useful for treating an
CC   inflammatory disorder, disorder of immune system and cancer selected
CC   from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
CC   viral disorder, in particular HIV infection and for modulating the
CC   growth of blood vessels. The polypeptides are useful for treating a
CC   tumour necrosis factor (TNF) mediated disorder, such as those selected
CC   from arthritis, inflammatory bowel disease and cardiac failure and a
CC   fibroblast growth factor-mediated disorder. It is also useful in assays
CC   to determine biological activity, to raise antibodies to isolate
CC   corresponding ligands or receptors, to quantify levels of protein or
CC   corresponding biological activity and as antiinflammatory agents,
CC   cognate corresponding ligand or receptors, as antiinflammatory agents,

Qy   400 SKWKTYKGIVNNEEKVFQGNSNFRDPVQNNNFIPPIVARVRVVPQTWHQRTALKVELIGC 459
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 skwkykgivnneekvfqgnsnfrdpvqnnifippivaryrvvrpqtwhqrtalkvelligc 360

Qy   460 QITQGNDSLVWRKTSOSTSVSTVKSKEDEITTRPIIPSEETSTGINITTVAIPLVLIVVLVFA 519
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 qitqgndslvwrktsqstsvstvktkedeitrpipseetstginitvaiplvllvivlvfa 420

Qy   520 GMGIFAFKRKKKKSSSPYGSAEAQKIDCWKQIKYPPARHIQOSAEFTISYDNEKEMYQKIDIL 579
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 gmgifafrkkkkgspygsaeaqktdcwkqikypfarhqsaeftisydnekemtqkidl 480

Qy   580 ITSDMAG 586
             |||||||
Db   481 itsdmag 487

CC   and in compositions for the treatment of skin, connective tissue and
CC   immune system diseases. The polynucleotide is useful as marker for
CC   tissue, as a chromosome marker or tags in the identification of a
CC   genetic disorder.
XX
SQ   Sequence 503 AA;

Query Match   42.6%; Score 1301.5; DB 21; Length 503;
Best Local Similarity   50.9%; Pred. No. 6.9e-108;
Matches 267; Conservative 26; Mismatches 33; Indels 199; Gaps 1;

Qy   61 AAGRGLLALLLAVSAAPLRLQAEELGLDGCGHLVTYQDSGTWTSKNXPGTYKPNHTVCEKTIT 120
              |||||||||:|:||:||||||||||:|||:|||||||||:|||||||||:|||:||||:|||:|-
Db   5 aggpsvlallfavcaplrlqaeelqgdgcghivtsqdsgtmtsknypgtypnytvcekiit 64

Qy   121 VPKGRKLILRGLDLDIESQTCASDYLLFTSSDQVGCPVCGSHMYVPKELLLNTSEVTVRRFE 180
              ||||:||||||||:||||:||||||||||||:||||:|||||||:|||||:|||:|||||:|
Db   65 vpkgkrliligdtlineskticasdyllfssatdqygvpicgswavpkelrllnsnevtvlfk 124

Qy   181 SCGSHISGRGFLLTYASSDDHPDLITCLERASHLKTEYSKFCPAGCRDVAGDIGSGNAMDGY 240
              |||||||||||||||||||:|||||||||||||:||||||||||||||:|:||||||:||||
Db   125 sgshisgrgflltyassdhpdliclergshyfeekysktcpagcrdiardisgntkdgy 184

Qy   241 RDTSLLCKAAIHAGITADELGQOISVLQRKGISRYEGTILANGVLSRDCGSLSDKRFLFTSN 300
              ||||||||||||||||:||||||||||||||||:||||||||||||||:|||||||||||
Db   185 rdtsllckaaihagitdelgghinillqskgishyeglangvlsrhgslsekrflf--- 241

Qy   301 GCSRSLSFEPDGQIRASSWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSNNHKPREW 360

Db   242 --- 241

Qy   361 LEIDLGEKRKKIIGIRTTGSTQSNFNFYVKSFVMNFKNNNSKWKTYKGIVNNEEKVFQGMS 420

Db   242 --- 241

Qy   421 NFERDPVQONNFIPIVARYVRVVPQTWHQRIALKVELIGCQIITQGNDSLVWRKTSQSTSVS 480

Db   242 --- 241

Qy   481 TKKEDETITRPIPSEEISTGINITTVAIPLVLLVVLVFAGMGIFAAFRKKKKSGPYGSA 540
              :|||:||||||:|||:|||||||||||||:||:||||||||:|||:||||||:||||
Db   242 -------tlpgmittvaipsviflailitgmgjfaictkrkkkgnpyvsa 285

Qy   541 EAQKIDCWKQIKPFARHQSAEFTISYDNEKEMTQKDLDLITSDMA 585
              ||||:||||||||||||||:|||||||||||||||:|||||||||
Db   286 daqktgcwkqikypfarhqsteftisydnekemtqkidlitsdma 330

---

us-09-691-344a-4.rag

RESULT 6
AAG75450

CC   AAH32943 to AAH37195 and AAG77514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression in a patient by rectifying
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the inactive proteins, diagnosis
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH31196 to AAH37204
CC   and AAG77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 583 AA;

Query Match   29.6%; Score 903.5; DB 22; Length 583;
Best Local Similarity   38.3%; Pred. No. 4.8e-72;
Matches 212; Conservative 92; Mismatches 183; Indels 67; Gaps 14;

Qy   43 APSGCVMYPGARGGGALARRAGRGL---LALLLAVSAPLRLQAEELGDGCGHLVTYQDSGT 99
              |:||||:||:|||::|||::||||---|:||||||:||||||||||:|||||||||||||
Db   2 apwaaiplsrslppcsnsssfsmplfllllvlvllledagagqdgcghtvlgpesgt 61

Qy   100 MTSKNYPGTYPNHTVCEKTITVPKGKRLILRLGDLDIE-SQTCASDYLLFTS---SSDQ 154
              :||::|||:||:|:||:|:|:|::|:|||:||:|:|:|-|:||:||:||:||---||:|
Db   62 ltsiuypqtypnstvcewelrivahmgcrvrikfqdtdedsdchfnylringigvsrte 121

Qy   155 YGPYCG--SMTVPKELLLNTSEVTVRFESGSHISGRGFLLTYASSDHPDLITCLERASHLI 213
              :|:||:--|::|:|:::|::||:|::||:||||||||||:||:|:||:|||:|||:|||:|
Db   122 igkycglqlqmnhsleskqneitllmsgihvsgrgflasysvidqdllccldtasnfl 181

Qy   214 KTEYSKFCPAGCRDVAGDIGGNMVDGYRDTSLLCKAAIHAGIIADELGQOISVLQRKGIS 273
              |:|:||||||||:|:|:|||:|:|:||||||||||||||||::|:||||||||:|:||:|
Db   182 epefskycpagclpfaeisgtlphgycdsspalcmagvhagvvsntlggqisvviskgip 241

Qy   274 RYEGTILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE---PDGQIRASSSWQSVNESGDQ 329
              |||:|||||||||::|||||:||||:|:||:|:||||---||||||||:|||||||:||
Db   242 yyesslannvtsvvghlstslftfktsgcygtlgmesgvriadpqitassvlewtdhtgge 301

Qy   330 VHWSPGQARLQDQGPSWASGDSSNNHKPRFWLEIDLGEKRKKIITGIRTTGSTQSNFNFYYK 389
              ::|:|::||||::|||:|:|||:|:|:|:||:|:||:|:|:|||:||||||:|||||||||
Db   302 nswkpkkarlkkppppwaafatde----yqwiqidinkekilgiltgltmvedhnyyvs 357

Qy   390 SFVMNFKNNNSKWKTYKGIVNNEEKVFQGSMSNRDPVQONNFIPIVARYVRVVPQTWHOR 449
              :::|::|||||:|:|:|:||::|:||||||:|:|||||:||||||||||||||||||:||:|
Db   358 ayrilysdddqkwtvyrepqveqdkifggnkdyhqdvrnxifplilarflirvnptqwqqk 417

Qy   450 IALKVELICCQ----------------ITQ----GNDSLVWRKTSQSISVSVSTKKEDETITRPI-- 492
              |:||:|||||----------------|||----||||||||||:|||::|||:||:||:|||||:--
Db   418 iamkmellgcqfqkprppkitgppprpnsnd----lkntappkiakgrapxftcqplqp 473

Qy   493 ------PSEET--STGINITT----------VAIPLVLVLLVFAGMGIFAAF------ 527
              ------||:||--||||||||----------||||||||:||||||||||||------
Db   474 ...

SEQ ID NO: 4
Database: A-Geneseq-032802
Accession No. AAB24216

RESULT 7

```
Qy   474  rssnefpaqtegqttaspdirnttvtpnvrtkdvalaavivpvlvmylttliilllvcawhvr  533
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   528  -RKKKKKGS---PY  537
          |||||||||| ||
Db   534  nrkkktegtydlpy  547
```

```
Db   AAY70539
ID   AAY70539 standard; protein; 669 AA.
XX
AC   AAY70539;
XX
DT   04-JUL-2000 (first entry)
XX
DE   Human Factor 8 Homologue.
XX
KW   Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW   cerebroprotective; therapeutic; coagulation related disorder;
KW   haemophilia; stroke; screening.
XX
OS   Homo sapiens.
XX
PN   WO200012532-A1.
XX
PD   09-MAR-2000.
XX
PF   20-AUG-1999;  99WO-US19047.
XX
PR   31-AUG-1998;  98US-0098521.
XX
PA   (ELIL) LILLY & CO ELI.
XX
PI   Rosteck PRJ, Su W, Li XM;
XX
DR   WPI: 2000-256580/22.
DR   N-PSDB; AAZ51872.
XX
PS   Claim 3; Page 64-66; 68pp; English.
XX
```

CC The present sequence is a human Factor 8 homologue (F8H),
CC a coagulation cofactor which is selectively expressed in
CC haematopoietic, heart and reproductive tissues. It has haemostatic and
CC cerebroprotective activities. The F8H contains a Factor 5/8 signature
CC and is useful as a therapeutic for treating coagulation related diseases
CC such as haemophilia and stroke. The nucleic acid is useful as
CC hybridisation probe and amplification primer for detecting deficiencies
CC in the level of F8H mRNA, for screening F8H gene mutations and for
CC monitoring regulation of gene expression. Fragments of the nucleic acid
CC are also useful as diagnostic probes and primers, and can be used in
CC screening methods such as those using DNA chips. The present sequence is
CC also useful as a target to screen therapeutically useful modulators
CC of the F8H.
XX
SQ Sequence 669 AA;

Factor 8 homolog polypeptides and nucleic acids encoding them, for
treating coagulation related disorders such as hemophilia and stroke -

```
Query Match             25.8%;  Score 786.5;  DB 21;  Length 669;
Best Local Similarity   35.9%;  Pred. No. 1.9e-61;
Matches  190;  Conservative  94;  Mismatches  164;  Indels  81;  Gaps  16;

Qy   124  GKRLLIRLGDLLDIE-SQTCASDYLLFYS----SSDQYGPYCG-SMYVPKELLLNTSEVTV  177
          |:|:  ||:||   | :||  |:|||:|     :|:| | ||  | :||||:|:|| ::|
Db     2  gervrikfgdtdledsdschfnylriyndigvsrtelgkycglglqmmhsleskgneiti  61

Qy   178  RFESGSHISGRGFLLTYASSDHHPDLTCLEKASHYLKTEYSKRPCAGCRDVACDISGNMV  237
          ||  |:| | || |: | | |  ||  | | ||:|  | | ||  || |  || |||:|
Db    62  lfmsjlhvsqrgflasysvidkqlltcldtasnflepefskycpagclipfaeisgtip  121

Qy   238  DGYRDTSLLCKMAIHAGIIADELGGQISVLQRKGISRYEGIIANGVLSRDGSLSDKRFLF  297
          |||:|||:||||:|||||||  | | |:|| ||||||  |||||||:|||| |:|| |||
```

RESULT 8

```
Db   AAB24216
ID   AAB24216 standard; protein; 889 AA.
XX
AC   AAB24216;
XX
DT   06-FEB-2001 (first entry)
XX
DE   Soluble neuropilin SNP-2 protein sequence SEQ ID NO:4.
XX
KW   Mouse; soluble neuropilin.
XX
OS   Mus sp.
OS   Synthetic.
XX
PN   JP2000236879-A.
XX
PD   05-SEP-2000.
XX
PF   17-FEB-1999;  99JP-0038920.
XX
PR   17-FEB-1999;  99JP-0038920.
XX
PA   (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR   WPI: 2000-658506/64.
DR   N-PSDB; AAC60787.
XX
PS   Claim 5; Page 13-17; 18pp; Japanese.
XX
```

CC The present invention describes a recombinant soluble neuropilin (I)
CC comprising an amino acid sequence in which the cell membrane-penetrating
CC region and/or the intracellular region is deleted from a mouse-derived
CC neuropilin. The soluble neuropilin can be used for the elucidation of
CC human bioreactions and disease conditions. The present sequence
CC represents a specifically claimed soluble neuropilin protein
CC sequence from the present invention.
XX
SQ Sequence 889 AA;

Soluble neuropilin having a deleted cell membrane-penetrating region,
used to diagnose human disease conditions -

```
Query Match             11.6%;  Score 353.5;  DB 21;  Length 889;

Qy   122  hgyrdsspicmagvhagvvsntigggisrvilskgiyyessiannvtsvvghplstslftl  181
          ||:  | | || :    :|       |:|  :  : |:||  || : | |  || | | ||
Db    122  TSNGCSRSLSFE-----PDGQIRASSWSVSNESGDQVHPSPGQARLQDQGPSWASGDSSN  353

Qy   182  ktsgcygtlgmesgviadpgltassvlewtdhtsqenswkpkarilkkppgpwaafatde  241
          |   ||  | | || | | :   :  | | |  ||:  :  : | :| |:  | |  |:|
Db    298  kifgnkdyhqdvrnifipplarflrvnpiqwqpkiankmelqscfipkxqrppkltqp  357

Qy   242  ---yqwiqidinkekkitgliittgsmvehnyyvsayrilysddgdqkwtvyrepgvead  297
          |  |  ||  || || ||| || ||  | :||  : |||  :|| |  || | || | |
Db    358  ppprnsnd---lkntappklakgrapkfcqplqprssnefpaqteqttaspdirnttv  413

Qy   354  NHKPREWLEIDLGEKKKKITGIRTTCSTQSNFNFYVKSFVMMFRKNNNSKWKTYKGIVNNEES  413
          |:| | | |     |:||||||   |::| |  :|  : |   | || :: :|  ||:|
Db    414  KVFQGNSNFRDPVQNNTFIPPIVARVRVRVPQTHQRIALKVELIGCQ--------ITQ-  463

Qy   414  tpvtkdvalaavivpvlvmylttliilllvcawhwrnrkkktegtydlpyw-----dragw  469
          |  ||||| ||:||| | || |||:||||||:||  |  : | |  ||   |     |:|
Db    470  wkgmkqfipakavdheetpvrysse--vnhlsprevttvi.gadsaeya  516
```

```
Qy   548  WKQIKYPF-----------------------ARHQSAEFYISYDNEKEMTQKLDLIITSDMA  585
          |:| :||                        | |:||  :|:|:| |  |   |: | |
Db    414  tpvtkdvalaavivpvlvmylttliilllvcawhwrnrkkktegtydlpyw-----dragw  469
```

```
Qy   464  --------GNDSLVWRKTRSQSTSVSTKKEDEtlTRPI-------PSEEI--STGINITY  505
Qy   414  KVFQGNSNFRDPVQNNTFIPPIVARVRVRVPQTHQRIALKVELIGCQ--------ITQ-  463
Qy   298  kifgnkdyhqdvrnifipplarflrvnpiqwqpkiankmelqscfipkxqrppkltqp  357
Qy   358  ppprnsnd---lkntappklakgrapkfcqplqprssnefpaqteqttaspdirntv  413
Qy   506  --------VAIPLVLLVLVFAGMGIFAAF---------RKKKKKGS---PYGSAEAQKTMDC  547
Qy   470  wkgmkqfipakavdheetpvrysse--vnhlsprevttvi.gadsaeya  516
```

Query Match   238   DGYRDTSLLCKMAIHAGIIADELGGQISVLQRKGISRYEGIIANGVLSRDGSLSDKRFLF 297

RESULT 9

ID   AAW96308 standard; Protein; 925 AA.
XX
AC   AAW96308;
XX
DT   28-JUN-1999　　(first entry)
XX
DE   Neuropilin-2.
XX
KW   KNNNSKKHKYKYFQGKNSNFRDPVQNNFIPIVARVYRVRVPGTWHQRIALKVE
KW   viral disease; oncological disease; cancer; screening; probes.
KW   viral pathogenesis; treatment; disease; graft rejection;
KW   neurone; development; cell growth; immune response;
KW   Neuropilin; neuropilin-2; semaphorin; transmembrane protein; axon;
XX
OS   Rattus rattus.
XX
PN   WO9904263-A1.
XX
PD   28-JAN-1999.
XX
PF   17-JUL-1998;　98WO-US14632.
XX
PR   17-JUL-1997;　97US-0052762.
XX
PA   (UUIO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Ginty DD, Kolodkin AL;
XX
DR   WPI; 1999-132446/11.
DR   N-PSDB; AAX08417.
XX
PT   New isolated semaphorin receptor, neuropilin-2 - used to develop

Best Local Similarity　23.1%; Pred. No. 2.3e-22;
Matches 120; Conservative 73; Mismatches 189; Indels 137; Gaps 17;

Qy   49 VPGARGGGALARAAGRGLLALLLAVSAPILRQAAELGD-GCGHLVTYQDSGTWTSKNYPG 107
               ::||:|:| :::   :|:|: | : |||: |:: :
Db   162 MTVPKELLLNTSEVTVRFESGGHISGRGPLLTV--------ASSD-------------   198

Qy   15 vpgstgdaa-------------------qparravrslqdpcgrpniskdagyitspsypq   57
               :: |:   :::| |: :| :||:||: |: | ::
Db   118 l-apptiissgsvliykftsdyargaqafslryeifktgsedcsknftspngtiespgfp   176

Qy   108 TYPNHFVCEKTITVPK-GRRILIRLG-DLDIESQYCASDYLLF------TSSSDQYGPYCGS 161
              |::|: |: |::| |: |:|| |: :: :| || ||||| ||::||| ||:|
Db   58 dyspbqncewlvyapepnqklvlnfnphfelekhdckydfieirdgdsesadllgkhcgn 117

Qy   162 MTVPKELLLNTSEVTVRFESGGHISGRGPLLTV--------ASSD-------------   198
            ::| || |::::| | :|:|| ||:||: ||:| :: :
Db   238 DGYRDTSLLCKAAIHAGIIADELGGGOISVLQRKGISRY----------------------EG 277

Qy   199 --HPDLITC-----LERASHYLKTEYSKF------------CPAGCRDVAGDIS--GNMV 237
              || | | |: |:||:: | : : | |::| | |:: || ::|
Db   177 ekyphnldcttllakprmeliligltfdlehdplqvgegdckydwldwdgiphvgpli 236

Qy   238 DGYRDTSLLCKAAIHAGIIADELGGGOISVLQRKGISRY----------------------EG 277
            |: |: || ||| ||| |:||| ||::|:| :|:|:| |
Db   237 gkygcgtktpskirsstgilsltfhtdmavakdgfsaryyllhqeppenfqcnvplgmesg 296

Qy   278 ILANGVLSRDGSLSDKRFLPTSNGCSRSLSFEPDGOIRASSSWQSVNESGDQVHWSPGQA 337
            ||:|| |:||:||||:|||:|||||||||||||||||||:|||:||:|||:|||:| |||
Db   237 gkygcgtktpskirsstgilsltfhtdmavakdgfsaryyllhqeppenfqcnvplgmesg 296

Qy   297 rianeqlsasstfsdgr--------------------wtpqgs 319
            :|:||:|:|||:|:||| ::|
Db   237 gkygcgtktpskirsstgilsltfhtdmavakdgfsaryyllhqeppenfqcnvplgmesg 296

Qy   338 RLQDOGPSNASGDSSNNHKPREWLEIDLGEKKKITGIRRTTG--STQSNFNPFYVKSFVMNF 395
            |||||||||||||||||||||||||||||||||||||||:|| |||||||:|:||:||||
Db   199 --HPDLITC-----LERASHYLKTEYSKF---------------CPAGCRDVAGDIS--GNMV 237

Qy   396 KNNNSKKHKYKYFQGKNSNFRDPVQNNFIPIVARVYRVRVPGTWHQRIALKVE 455
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   376 stngedvmvyrhgknh--klfqanndatevvlokl hmpliterfilrirpqtwhlgialrie 433

Qy   456 LIGCQLTQGNDSLVWRKTS-------OGTRSVSTRKEDETTTRP 491
            ||||||||||||||||||| |||||||||||||||||
Db   434 lfgcrvtdapcsnmlgmlsqliadtcqisasstreyiwsp 472

Sequence　925 AA;

RESULT 10

ID   AAB64627 standard; Protein; 439 AA.
XX
AC   AAB64627;
XX
DT   22-MAR-2001　　(first entry)
XX
DE   Human secreted protein BLAST search protein SEQ ID NO: 137.
XX

Best Local Similarity　11.5%;　Score 351.5; DB 20;　Length 925;
Matches 114; Conservative 76; Mismatches 181; Indels 97; Gaps 17;

Qy   88 CGHLVTYQDSGTWTSKNYPGTYPNHFVCEKTITVPK-GRRILIRLG-DLDIESQYCASDY 145
              | ||:|||||||||||||||||||||||||||| ||:||:|| :: :|| || ||
Db   28 cggrlnskdagyitspsypqdyspbqncewlvyapepnqklvlnfnphfelekhdckydf   87

Qy   146 LLF--TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGGHISGRGPLLTV----AS 196
            ||: ||::||| ||:|:::| || |::::| | :|:|| ||:||: ||:| |
Db   88 ieirdgddsesadllgkhcgni--apptiissgsvliykftsdyargaqafslryeiftgs 146

Qy   197 SD-------HPDLITC-----LERASHYLKTEYSKF---------LERASHYLKTEYSKF 220
            || || | | |: |:||:: | : : |:||:: | : :
Db   88 ekyphnldcttllakprmeliligltfdlehdplqvgeg 206

Qy   221 -CPAGCRDVAGDIS---GNMVDGYRDTSLLCKAAIHAGIIADELGGGOISVLQRKGISRY-- 275
             | |::| || ::| ||::|: |: || ||| ||| |:||| ||::|:| :|:|:|
Db   147 edcsknftspngtiespgfpkdkydwldwdgiphvgpli 206

Qy   147 edcsknftspngtiespgfpkdkydwldwdgiphvgpli 206
            : | : | ::| || ::|
Db   207 dckydwldwdgiphvgpligkyogtktpskirsstgilsltfhtdmavakdgfsaryyl 266

Qy   276 vhqeppenfqcnaplgmesgriane 300
            |:||||||||| ||:|||:|||:||
Db   207 dckydwldwdgiphvgpligkyogtktpskirsstgilsltfhtdmavakdgfsaryyl 266

Qy   267 vhqeppenfqcnaplgmesgriane 291
            ||:||||||||| ||:|||:|||:|
Db   207 dckydwldwdgiphvgpligkyogtktpskirsstgilsltfhtdmavakdgfsaryyl 266

Qy   329 QVHWSPGQARLQDOGPSNASGDSSNNHKPREWLEIDLGEKKKITGIRRTTG--STQSNFNP 386
            |||||||||||||||||||||||||||||||||||||||||||||||:|| ||||||||
Db   301 dgrwtpqgsribgddanqgvtpuvdsn----keylqvdlrfltmlalatcqaisretqkgy 356

Qy   387 YVKSFVMNFKNNNSKKHKYKYFQGKNSNFRDPVQNNFIPIVARVYRVRVPGTW 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 yvksyklevstngedvmvyrhgknh--kvfqanndatevvlokl hmpliterfilrirpqtw 414

Qy   447 HQRIALKVELIGCQLTQGNDSLVWRKTS-------OGTRSVSTRKEDETTTRP 491
            ||||||||||||||||||||||||||| |||||||||||||||||
Db   415 higiallelfgcrvtdapcsnmlgmlsqliadtcqisasstreyiwsp 462

Sequence　925 AA;

PT   products for the diagnosis and treatment of neurological,
PT   immunological, oncological and viral diseases
XX
PS
CC   The neuropilins are type I transmembrane proteins and act as
CC   semaphorin III (Sema III) receptors. The semaphorins have been
CC   shown to function in repulsive axon guidance. Sema III is a
CC   secreted protein that in vitro causes neuronal growth cone
CC   collapse and chemorepulsion on neurites and is required in vivo for
CC   correct sensory afferent innervation and other aspects of
CC   development. Agents which inhibit or enhance the interaction of a
CC   semaphorin and a neuropilin can be potent modulators of nerve cell
CC   growth, immune responsiveness, and viral pathogenesis, and can be
CC   used in the treatment and diagnosis of neurological disease
CC   neuro-regeneration, and diagnosis of neurological disease
CC   including hypersensitivity and
CC   infection/diseases. The neuropilins, neuropilin-encoding nucleic
CC   acids, and unique portions also are useful in screening chemical
CC   libraries for regulators of semaphorin-mediated cell activity, and in
CC   genetic mapping as probes for related genes, as diagnostic reagents
CC   for genetic, neurological and oncological disease.
XX
SQ   Sequence 925 AA;

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
OS Homo sapiens.
PI Rosen CA; Ruben SM; Komatsoulis GA;
PN WO200077197-A1.
PR 11-JUN-1999.
PD 21-DEC-2000.
PF 01-JUN-2000; 2000WO-US14934.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
DT WPI; 2001-033212/04.

Disclosure; Page 524-526; 558pp; English.

CC Isolated nucleic acid molecule encoding a human secreted protein is
CC used in preventing, treating or ameliorating a medical condition -
CC The invention relates to the isolation of genes AAF32757-F32803 encoding
CC the human secreted proteins SEQ. The sequence is a search
CC result from a BLASTX homology search. The genes and proteins are useful
CC for preventing, ameliorating or treating medical conditions, e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer, and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.

PN Sequence   439 AA;

Query Match   11.4%;   Score 349; DB 22;   Length 439;
Best Local Similarity   25.3%;   Pred. No. 1.9e-22;
Matches   105;   Conservative   72;   Mismatches   166;   Indels   84;   Gaps   16;

Qy   88   CGHLVTYQDSGTMTSKNYPCGTYPNHTVCEKTTTVPK-GKRLLRLG-DLDIESOTCASDY 145
          | |::| | | :| ::||| |:|::|| :::| | ::| |:: |::| ::|: | ||
Db   28   CGGRInSKdagylspydPdDYdpshqncewivyapepnqkivlnfnphfelekhdckydf 87

Qy   146 LLF-----TSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFFLTY----AS 196
          :|| | :| ||||| ::| ||:| ::||| |||:::| ::|| || | |:|
Db   88   leirdgdsesadligkhcqni-apptissgsmlyikftsdyargagfslryeifktgs 146

Qy   197 SD----------------------HPDLITC----LERASHYLKTEYSKF--------- 220
          || ||| :|| |::|::| :|||:|
Db   147 edosknfspngtlespgffpekyphnldctftilakpkmeillqfilfdlehdplgveg 206

Qy   221 -CPAGCROVAGDIS--GNMVDGYRDTSLCLKAAIHAGIIADELGGOISVLQRKGISRYEG 277
          |||| ::| ||:| |:::| :|:|| | :||| :|| || :|||:| ||:| |::||||
Db   207 dckydkdldivdgiphvgplgkyeqtktpselrsstgl:slthtdmavak------dg 259

Qy   278 ILANGVTLSRDGSLSSDKRFLFTSNGCSRSLSFE-----PDGGIRASSSMGQSVNESGDQVHMS 333
          |||::|:|| |:: |::|| |:||| || :| | || :||::::|||:| ||| |:
Db   260 fsaryylvhqeplenfq-----------cnvplgmesgrlaneqisassty------sdgrvt 305

Qy   334 PGQARLQDQGFPSWAGSGDSSNNHKPREWLEIDLGEKKIIGIRTTG--STQSNFNPTYVKSF 391
          |:::|::|:: | | :|: || ::::| ::::|:: ::::|::| | :| | :|:| |
Db   306 pqgsllbgddngvtgnld5n----keylqvdlrfltmltalataggalsretgngyyvksy 361

Qy   392 VNNFKNNNSKWKTYKGIVNNEEKVFQGJNSNFRDPVQVNNEFIPPIVARVRVRVPPQTWHQRIA 451
          ||:|:::::::|: :|:: :: ::| | |: :|:::: ::::|:|:||:| ::|:|| | |
Db   362 klevstngedwmvyrhgknh--kvfganndatevvlnklhapltfflvrlrpqtwhegla 419

Qy   452 LKVELICCOTT 462
          |::| || |||
Db   420 lrlelfccrvt 430

RESULT 11
AAY06320
ID   AAY06320 standard; Protein; 555 AA.
AC   AAY06320;
DT   06-SEP-1999 (first entry)
DE   Human soluble neuropilin-2.
OS   Homo sapiens.
OS   Synthetic.
KW   WO929858-A1.
KW   Neuropilin-2; human; vascular endothelial growth factor receptor;
KW   VEGF165R; inhibitor; metastasis; angiogenesis; inflammation;
KW   arthritis; diabetic retinopathy; cytostatic; antiinflammatory;
KW   antiangiogenic; antidiabetic; ophthalmological; antiarthritic;
KW   soluble.
PN   09-DEC-1998.
PR   09-SEP-1998; 98US-0099615.
PR   09-DEC-1997; 97US-0069155.
PR   12-DEC-1997; 97US-0069687.
PA   (CHIL-) CHILDRENS MEDICAL CENT.
PI   Gagnon ML, Klagsbrun M, Soker S;
DT   WPI; 1999-394975/33.
DR   N-PSDB; AAX59154.

Claim 6; Page 90-91; 96pp; English.

Soluble human neuropilins and related polynucleotides

CC   The present sequence represents a soluble, truncated form of human
CC   neuropilin-2, a novel receptor of vascular endothelial growth
CC   factor (VEGF). The soluble neuropilin-2 deviates from the
CC   full-length protein (see AAY06318) between the b2 and c domains
CC   after amino acid 547, at the position of an exon-exon boundary;
CC   and includes 8 extra c-terminal amino acids. Soluble neuropilins
CC   are capable of binding to VEGF165 or any form of VEGF that contains
CC   exon 7 and are useful for inhibiting VEGF interaction not only
CC   with neuropilins but also with KDR/Flk-1 and Flt1. They may also
CC   act as dominant negative receptors when expressed in cells by
CC   dimerizing with intact neuropilin receptors. The soluble
CC   neuropilins are useful for treatment of diseases or disorders
CC   associated with VEGF, e.g. metastasis, inappropriate angiogenesis,
CC   chronic inflammation, diabetic retinopathy and arthritis. Solid
CC   tumors can also be treated.

SQ   Sequence   555 AA;

RESULT 12

```
AAW96254
ID   AAW96254 standard; Protein; 909 AA.
XX
AC   AAW96254;
XX
DT   12-MAY-1999  (first entry)
XX
DE   Mouse semaphorin receptor SR2(b) 0.
XX
KW   Semaphorin receptor; SR; regulation; cell function;
KW   cell morphology; nerve cell; neuron; axon; diagnosis; therapy;
KW   detection; mutant allele; neuropilin.
XX
OS   Mus musculus.
XX
XX
PN   WO9902556-A1.
XX
PF   21-JAN-1999.
XX
PR   08-JUL-1998;  98WO-US14290.
XX
PR   24-SEP-1997;  97US-0936135.
PR   08-JUL-1997;  97US-0889458.
XX
PA   (RECC ) UNIV CALIFORNIA.
XX
PI   Chen H, He Z, Tessier-Lavigne M;
XX
DR   WPI; 1999-120780/10.
DR   N-PSDB; AAX09003.
XX
PT   New semaphorin receptor (SR) class 1 and class 2 proteins - useful
PT   for, e.g. regulating nerve cell function and morphology
XX
PS   Claim 1; Page 84-86; 105pp; English.
XX
```

Best Local Similarity 25.3%; Pred. No. 2.7e-22; Indels 84; Gaps 16;
Matches 109; Conservative 72; Mismatches 166;

```
Qy    88 CGHLVTYQDSGTMYSKRNYPGTYPNHTVCEKTITYPK-GKRLILRLG-DLDIESQYCASDY 145
            :||  :|||:| |::  ::| :|  ||  :||:|:: | |::|:|:| |: :: : | :
Db    28 cggrlnskdagyltspsypyqpdypshqncewivyapepnqkivlnfnphfelekhdckydf 87

Qy   146 LIF-----TSSSDQYGPYCGSSMYVPKELLLNTSEYTVRFESGSHISGRGFLLIY-----AS 196
            |:|       |::|:| ||::| :||||:|  | | ||:  |:||  |  ||:       :
Db    88 leirdgdseadiglqkhcgni-apptlissgsmjykftsdyarggaqfslryelfktcgs 146

Qy   197 SD-----------HPDLITC----LERASHYLKTEYSKF----------------------- 220
            |:                    |:  |: |:| |
Db   147 edcskhftspngtiespgffpekyphnidctftilakpkmeilqlfilidlehdplqvgeg 206

Qy   221 -CPAGCRDVAGDIS--GNMVDGYRDYSLLCKAAIHAGIIADELGGOISVLQRKGISRYEG 277
            |  |  | | :|     ||:|  | :: |  | |:  :  :| |:|   | |: |: |
Db   207 dckydwldiwdgiplhvgpligkycgtktpselrsstglislfthtdmavak------dg 259

Qy   278 ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE------PDGQIRASSSMQSVNESGDQVHWS 333
            ||| |:|::|| :  |: |: |   |  |          | ::|:   |:|::  | : |::
Db   260 fsaryylvhqeplenfg-----------cnvplgmesgrlaneqiaassty----sdgrwt 305

Qy   334 PCGARLQDOGPSWASGCDSSNNHKPREWLEIDLGEKKKITGIRTTG--STOSNFYVKSF 391
            |  | :||: ||| |  :| |:  |:  |:| : : |  |: | |    | ||: | |:
Db   306 pqgsrlhgddngvtcpldsn---keylqydlrfltmltaiatcqgaisretcqngyyvksy 361

Qy   392 VMNFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVRVPQTWHQRIA 451
            |:  :| | |   |:|:   :: : |   :   ||| :  :   | ||  ||||   | |
Db   362 klevstngedwmyrighknh-kvfganndatevvlnklhapllrfvrirpqtwhsgia 419

Qy   452 LKVELLGCQIP 462
            |:|||:||||
Db   420 lrlelfgcvrt 430
```

```
CC   Semaphorin receptor (SR) proteins are involved in nerve cell
CC   function and morphology. The SR proteins can regulate cell
CC   guidance. The SR proteins, e.g. nerve cells. Inhibitors of SR
CC   growth. Binding agents specific to the SR polypeptides are useful
CC   for diagnosis and therapy, especially where disease or disease
CC   prognosis is associated with improper or undesirable axon outgrowth
CC   or orientation. The SR nucleic acid sequences are used to provide
CC   translatable transcripts, hybridisation probes, PCR primers,
CC   diagnostic nucleic acids, etc. They are used in detecting the
CC   presence of SR genes, and to identify wild type and mutant alleles.
CC   Mutant alleles are used to generate allele-specific oligonucleotide
CC   probes for clinical diagnosis. The methods are used to identify
CC   agents which modulate the activity of SR polypeptides. This
CC   sequence encodes the mouse SR2(b) 0 polypeptide which is an isoform
CC   of neuropilin-2b with 0 amino acids inserted.
XX
SQ   Sequence 909 AA;
```

RESULT 13

```
AAW96255
ID   AAW96255 standard; Protein; 926 AA.
XX
AC   AAW96255;
XX
DT   12-MAY-1999  (first entry)
XX
DE   Human semaphorin receptor SR2a(0) 17.
XX
KW   Semaphorin receptor; SR; regulation; cell function;
KW   cell morphology; nerve cell; neuron; axon; diagnosis; therapy;
KW   detection; mutant allele; neuropilin.
XX
OS   Homo sapiens.
XX
XX
PN   WO9902556-A1.
```

Query Match 11.4%; Score 349; DB 20; Length 909;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 109; Conservative 72; Mismatches 166; Indels 84; Gaps 16;

```
Qy    88 CGHLVTYQDSGTMYSKRNYPGTYPNHTVCEKTITYPK-GKRLILRLG-DLDIESQYCASDY 145
            :||  :|||:| |::  ::| :|  ||  :||:|:: | |::|:|:| |: :: : | :
Db    28 cggrlnskdagyltspsypyqpdypshqncewivyapepnqkivlnfnphfelekhdckydf 87

Qy   146 LIF-----TSSSDQYGPYCGSSMYVPKELLLNTSEYTVRFESGSHISGRGFLLIY-----AS 196
            |:|       |::|:| ||::| :||||:|  | | ||:  |:||  |  ||:       :
Db    88 leirdgdseadiglqkhcgni-apptlissgsmjykftsdyarggaqfslryelfktcgs 146

Qy   197 SD-----------HPDLITC----LERASHYLKTEYSKF----------------------- 220
            |:                    |:  |: |:| |
Db   147 edcskhftspngtiespgffpekyphnidctftilakpkmeilqlfilidlehdplqvgeg 206

Qy   221 -CPAGCRDVAGDIS--GNMVDGYRDYSLLCKAAIHAGIIADELGGOISVLQRKGISRYEG 277
            |  |  | | :|     ||:|  | :: |  | |:  :  :| |:|   | |: |: |
Db   207 dckydwldiwdgiplhvgpligkycgtktpselrsstglislfthtdmavak------dg 259

Qy   278 ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE------PDGQIRASSSMQSVNESGDQVHWS 333
            ||| |:|::|| :  |: |: |   |  |          | ::|:   |:|::  | : |::
Db   260 fsaryylvhqeplenfg-----------cnvplgmesgrlaneqiaassty----sdgrwt 305

Qy   334 PCGARLQDOGPSWASGCDSSNNHKPREWLEIDLGEKKKITGIRTTG--STOSNFYVKSF 391
            |  | :||: ||| |  :| |:  |:  |:| : : |  |: | |    | ||: | |:
Db   306 pqgsrlhgddngvtcpldsn---keylqydlrfltmltaiatcqgaisretcqngyyvksy 361

Qy   392 VMNFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVRVPQTWHQRIA 451
            |:  :| | |   |:|:   :: : |   :   ||| :  :   | ||  ||||   | |
Db   362 klevstngedwmyrighknh-kvfganndatevvlnklhapllrfvrirpqtwhsgia 419

Qy   452 LKVELLGCQIP 462
            |:|||:||||
Db   420 lrlelfgcvrt 430
```

```
XX
PD   21-JAN-1999.
XX
XX
PF   08-JUL-1998;  98WO-US14290.
PR   24-SEP-1997;  97US-0936135.
PR   08-JUL-1997;  97US-0889458.
XX
PA   (RRGC ) UNIV CALIFORNIA.
XX
PI   Chen H, He Z, Tessier-Lavigne M;
XX
DR   WPI; 1999-120780/10.
DR   N-PSDB; AAX09002.
XX
PT   New semaphorin receptor (SR) class 1 and class 2 proteins - useful
PT   for, e.g. regulating nerve cell function and morphology
XX
PS   Claim 1: Page 81-83; 105pp; English.
XX

CC   Semaphorin receptor (SR) proteins are involved in nerve cell
CC   guidance. The SR proteins can regulate cell function and morphology,
CC   especially nerve cells. Inhibitors of SR proteins promote nerve cell
CC   growth. Binding agents specific to the SR polypeptides are useful
CC   for diagnosis and therapy, especially where disease or disease
CC   prognosis is associated with improper or undesirable axon outgrowth
CC   or orientation. The SR nucleic acid sequences are used to provide
CC   translatable transcripts, hybridisation probes, PCR primers,
CC   diagnostic nucleic acids, etc. They are used in detecting the
CC   presence of SR genes, and to identify wild type and mutant alleles.
CC   Mutant alleles are used to generate allele-specific oligonucleotide
CC   probes for clinical diagnosis. The methods are used to identify
CC   agents which modulate the activity of SR polypeptides. This
CC   sequence is the human SR2a(0) 17 polypeptide which is an isoform of
CC   neuropilin-2a with 17 amino acids inserted after amino acid 809.
XX
SQ   Sequence 926 AA;
```

```
Query Match             11.48%;  Score 349;  DB 20;  Length 926;
Best Local Similarity   25.3%;   Pred. No. 6.2e-22;
Matches  109;  Conservative  72;  Mismatches 166;  Indels 84;  Gaps 16;

Qy         88 CGHLVTYQDSGTMTSKNTFGTYFNHTVCEKTITVFK-CKRLIRLG-DLDIESQTCASDY 145
                 |||   ||   ||||  | ||    || | ||| |||    |    |  |   | |
Db         28 cggrlnskdagyltspypqdypshqqocewivyapepnqkivlnfnphfelekhdcxydf 87

Qy        146 LLF-----TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY----AS 196
                 |          ||||     |  ||  || |  |   |||   | | |||     |
Db         88 leirdgdseaadligkhcgni-appltissgsmlyiktsdyarqgagfslryelrktgs 146

Qy        197 SD---------HPDLITC-----LERASHYLKTEYSKF------------------- 220
                 |               |      ||  ||| |  | ||
Db        147 edcsknftspgtlespgfpekyphnldctlfillakpkmeiilqlifildlehdglqyeq 206

Qy        221 -CPAGCRDVACDIS--GNMVDGYRDTSLJCKAAIHAGIIADELGGOISVLQRKGISRYEG 277
                |     |  ||     | |||  || |        |       |  | |  |  |||
Db        207 dckydwidivdgiphvgpligkycqtktpselrsstglisltfthdmavak----dg 259

Qy        278 ILANGVLSRDGSLSDKRFLFTSNGCGSRSLSFE----PDGQIRASSWQSVNESGDGVHWS 333
                |  |      ||   |   |      | |   |        |    |   |  ||
Db        260 fsaryylvhqeplenfq------civplgmesgrianegisassty----sdgrwt 305

Qy        334 PCQARLQDGCPSWASCDSSNNHKPREWLEIDLGEKKRKITTGITRTTG--STQOSNFNFYVKSF 391
                || ||  |      ||   | |   |  |  |   |||    | |      |   ||    | |
Db        306 pqgsrlhgddngwtpnldsn----keylqvdrifltmltaiatcgaisretqngyvyksy 361

Qy        392 VNMFKNNNSKKWKTYKGITVNNEEKVYPGQNSNFRDPVQONNFIPPIVARYVRVPWTFHQRIA 451
                   |           |   | | |    | |        |   |   |       |  |  |
Db        362 klevstngedvmvyrhgknh--kvfqaandatevvlnklhaplitrfvrirpqtwhsgla 419

Qy        452 LKVELLGCOIT 462
                :||:|| ||:|
Db            LKVELLGCOIT 462
```

```
Db            lrielgcrvt 430                                        420

RESULT 14
AAY06318
ID   AAY06318 standard; protein; 931 AA.
XX
AC   AAY06318;
XX
DT   06-SEP-1999 (first entry)
XX
DE   Human neuropilin-2.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FH
FT   Peptide        1..23
FT                  /note="signal peptide"
FT   Protein        24..931
FT                  /note="mature protein"
FT   Domain         24..148
FT                  /note="domain a1"
FT   Domain         149..275
FT                  /note="domain a2"
FT   Domain         277..433
FT                  /note="domain b1"
FT   Domain         434..594
FT                  /note="domain b2"
FT   Domain         642..800
FT                  /note="domain c"
FT   Domain         865..893
FT                  /note="transmembrane domain"
FT   Domain         8941..931
FT                  /note="cytoplasmic domain"
XX
PN   WO9929858-A1.
XX
PD   17-JUN-1999.
XX
PF   09-DEC-1998; 98WO-US26138.
PR   09-SEP-1998; 98US-0096615.
PR   09-DEC-1997; 97US-0069155.
PR   12-DEC-1997; 97US-0069687.
XX
PA   (CHIL-) CHILDRENS MEDICAL CENT.
XX
PI   Gagnon ML, Klagsbrun M, Soker S;
XX
DR   WPI; 1999-394975/33.
DR   N-PSDB; AAX59152.
XX
PT   Soluble human neuropilins and related polynucleotides
XX
PS   Claim 4: Page 86-87; 98pp; English.
XX

CC   The present sequence represents human neuropilin-2, a novel
CC   receptor of vascular endothelial growth factor (VEGF). Expression
CC   of neuropilin is associated with angiogenesis and motile
CC   metastatic cancer cells, and thus is an important target for
CC   antiangiogenic and anticancer therapy. Neuropilin-2 cDNA (see     A
CC   AAX59152) was obtained from MDA-MB-231 cells by expression cloning.
CC   soluble neuropilin (see AAY06320) comprising amino acids 277-593 of
CC   the present sequence is claimed, and can be isolated from neuropilin
CC   producing cells or from any form of VEGF that contains exon 7 and are
CC   encoding DNA. Soluble truncated neuropilins are capable of
CC   binding to VEGF165 or any form of VEGF interaction not only with
CC   therefore useful for inhibiting VEGF interaction not only with
```

Left column

```
CC   neuropilins but also with KDR/Flk-1 and Flt1. The soluble NPs may
CC   also act as dominant negative receptors when expressed in cells by
CC   dimerizing with intact neuropilin receptors. The soluble
CC   neuropilins are useful for treatment of diseases or disorders
CC   associated with VEGF, e.g. metastasis, inappropriate angiogenesis,
CC   chronic inflammation, diabetic retinopathy and arthritis. Solid
CC   tumors can also be treated.
XX
SQ   Sequence   931 AA;

Query Match       11.4%;   Score 349;   DB 20;   Length 931;
Best Local Similarity   25.3%;   Pred. No. 6.2e-22;
Matches 109;   Conservative   72;   Mismatches 166;   Indels   84;   Gaps   16;

Qy    88  CGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTIITVPK-GKRLILRLG-DLDIESQTCASDY 145
Db    28  cggrinskdagyitspsypdqypshqncewivyapepngkivlnfnphfeiekhdckydf   87

Qy   146  LLF----TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY-----AS 196
Db    88  leirdgdsesadilgkhcgni-appilssgsmiyikftsdyarggagfslryeiktgs  146

Qy   197  SD-----------------------HPDLITC-------LERASHYLKTEYSKF------ 220
Db   147  edcsknftspngtiespgfpekyphnldctftilakpkmeiliqfilfidlehdplqvgeg 206

Qy   221  -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRYEG 277
Db   207  dckydwidiwdgiphvgpligkycgtktpselrsstgjlsltfhtdmavak------dg  259

Qy   278  ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE-----PDGQIRASSSWQSVNESGDQVHWS 333
Db   260  fsaryylvhqeplenfq------cnvplgmesgrilaneqlsassty-----sdgrwt  305

Qy   334  PGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTTG--STQSNFNFYVKSF 391
Db   306  pqqsrlhgddangvtppldsn---keylqydlrflmltalatqgalsretqngyyvksy  361

Qy   392  VNNFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVPVPQTWHQRIA 451
Db   362  klevstngedwmvyrhgknh--kvfqanndatevvlnklhapliltrfvrirpqtwhsgia 419

Qy   452  LKVELGCQTT 462
Db   420  lrlelfgcrvt 430
```

```
RESULT 15
AAY14563
ID   AAY14563 standard; protein; 931 AA.
XX
AC   AAY14563;
XX
DT   14-SEP-1999   (first entry)
XX
DE   Human VEGF(165)R/NP-2 protein.
XX
OS   Homo sapiens.
XX
KW   Human; vascular endothelial growth factor; VEGF; receptor; neuropilin;
KW   antagonist; semaphorin; collapsin; cancer; metastasis; angiogenesis;
KW   diabetic retinopathy; osteoarthritis; psoriasis.
XX
PN   WO9929729-A2.
XX
PD   17-JUN-1999.
XX
PF   09-DEC-1998;   98WO-US26114.
XX
PR   19-MAR-1998;   98US-0078541.
PR   09-DEC-1997;   97US-0069155.
PR   12-DEC-1997;   97US-0069687.
```

Right column

```
(CHIL-) CHILDRENS MEDICAL CENT.
XX
PA   Klagsbrun M, Miao H, Soker S, Takashima S;
XX
PI   1999-399459/33.
XX
DR   WPI; 1999-399459/33.
DR   N-PSDB; AAX79350.
XX
PT   Neuropilin receptor antagonists is useful for treatment of, e.g.
PT   metastatic cancer.
XX
PS   Disclosure; Fig 12; 80pp; English.
XX
CC   This sequence represents the human VEGF(165)R/NP2 (or neuropilin-2).
CC   The invention relates to a neuropilin antagonist that binds neuropilin
CC   and has VEGF antagonistic activity as determined by the human umbilical
CC   vein endothelial cell (HUVEC) proliferation assay using VEGF165. The
CC   neuropilin antagonist (semaphorin/collapsin family member) is useful for
CC   treatment of a disease or disorder associated with VEGF, including cancer
CC   such as metastatic cancer and angiogenesis diseases, diabetic retinopathy,
CC   osteoarthritis and psoriasis.
XX
SQ   Sequence   931 AA;

Query Match       11.4%;   Score 349;   DB 20;   Length 931;
Best Local Similarity   25.3%;   Pred. No. 6.2e-22;
Matches 109;   Conservative   72;   Mismatches 166;   Indels   84;   Gaps   16;

Qy    88  CGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTIITVPK-GKRLILRLG-DLDIESQTCASDY 145
Db    28  cggrinskdagyitspsypdqypshqncewivyapepngkivlnfnphfeiekhdckydf   87

Qy   146  LLF----TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY-----AS 196
Db    88  leirdgdsesadilgkhcgni-appilssgsmiyikftsdyarggagfslryeiktgs  146

Qy   197  SD-----------------------HPDLITC-------LERASHYLKTEYSKF------ 220
Db   147  edcsknftspngtiespgfpekyphnldctftilakpkmeiliqfilfidlehdplqvgeg 206

Qy   221  -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRYEG 277
Db   207  dckydwidiwdgiphvgpligkycgtktpselrsstgjlsltfhtdmavak------dg  259

Qy   278  ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE-----PDGQIRASSSWQSVNESGDQVHWS 333
Db   260  fsaryylvhqeplenfq------cnvplgmesgrilaneqlsassty-----sdgrwt  305

Qy   334  PGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTTG--STQSNFNFYVKSF 391
Db   306  pqqsrlhgddangvtppldsn---keylqydlrflmltalatqgalsretqngyyvksy  361

Qy   392  VNNFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVPVPQTWHQRIA 451
Db   362  klevstngedwmvyrhgknh--kvfqanndatevvlnklhapliltrfvrirpqtwhsgia 419

Qy   452  LKVELGCQTT 462
Db   420  lrlelfgcrvt 430
```

Search completed: July 9, 2002, 02:13:59
Job time: 4006 sec

```
OM protein - protein search, using sw model

                              GenCore version 4.5
                    Copyright (c) 1993 - 2000 Compugen Ltd.

Run on:         July 9, 2002, 01:12:13 ; Search time 66.82 seconds
                                         (without alignments)
                                         214.208 Million cell updates/sec

Title:                US-09-691-344A-4
Perfect score:        3052
Sequence:             1 MGFGAQORLRPVPAPRSSAE........YDNEKEMTQKLDLITSDMAG 586

Scoring table:        BLOSUM62
                      Gapop 10.0 , Gapext 0.5

Searched:             231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters:    231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :       Issued_Patents_AA.*
              1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
              2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
              3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
              4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
              5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
              6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	349	11.4	909	3	US-08-936-135-18	Sequence 18, Appl
2	349	11.4	926	3	US-08-936-135-20	Sequence 20, Appl
3	343.5	11.3	901	3	US-08-936-135-22	Sequence 22, Appl
4	343.5	11.3	906	3	US-08-936-135-24	Sequence 24, Appl
5	343.5	11.3	909	3	US-08-936-135-8	Sequence 8, Appl
6	343.5	11.3	909	3	US-08-936-135-10	Sequence 10, Appl
7	343.5	11.3	914	3	US-08-936-135-12	Sequence 12, Appl
8	343.5	11.3	926	3	US-08-936-135-14	Sequence 14, Appl
9	343.5	11.3	931	3	US-08-936-135-16	Sequence 16, Appl
10	329	11.0	923	3	US-08-936-135-6	Sequence 6, Appl
11	321	10.5	218	3	US-07-607-538C-4	Sequence 4, Appl
12	321	10.5	218	3	US-08-162-402B-4	Sequence 4, Appl
13	318	10.4	2183	2	US-08-746-111-5	Sequence 5, Appl
14	311	10.2	160	2	US-08-162-402B-14	Sequence 14, Appl
15	306.5	10.0	2343	4	US-09-324-867-2	Sequence 2, Appl
16	299	9.8	217	1	US-07-607-538C-3	Sequence 3, Appl
17	299	9.8	217	1	US-08-162-402B-3	Sequence 3, Appl
18	299	9.8	217	2	US-07-607-538C-2	Sequence 2, Appl
19	299	9.8	217	2	US-08-162-402B-2	Sequence 2, Appl
20	299	9.8	218	2	US-08-162-402B-6	Sequence 6, Appl
21	299	9.8	387	2	US-08-162-402B-8	Sequence 8, Appl
22	293	9.6	465	2	US-08-162-402B-12	Sequence 12, Appl
23	293	9.6	159	2	US-08-162-402B-12	Patent.No.5422260
24	289.5	9.5	2351	4	US-09-203-916-1	Sequence 1, Appl
25	285.5	9.4	1438	4	US-08-683-839B-3	Sequence 3, Appl
26	285.5	9.4	1471	1	US-08-882-083-2	Sequence 2, Appl
27	285.5	9.4	1661	2	US-08-558-107-2	Sequence 2, Appl
28	285.5	9.4	1661	4	US-09-243-539-2	Sequence 2, Appl
29	285.5	9.4	2332	1	US-07-864-004B-4	Sequence 4, Appl
30	285.5	9.4	2332	1	US-08-251-937A-4	Sequence 4, Appl
31	285.5	9.4	2332	1	US-08-212-133A-2	Sequence 2, Appl
32	285.5	9.4	2332	1	US-08-276-594A-2	Sequence 2, Appl
33	285.5	9.4	2332	1	US-08-474-503-2	Sequence 2, Appl
34	285.5	9.4	2332	2	US-08-670-707A-2	Sequence 2, Appl
35	285.5	9.4	2332	4	US-09-037-660-2	Sequence 2, Appl
36	285.5	9.4	2332	3	US-09-324-867-3	Sequence 3, Appl
37	285.5	9.4	2351	6	US-US93-03275-4	Sequence 4, Appl
38	284.5	9.3	218	5	PCT-US94-13200-2	Sequence 2, Appl
39	285.5	9.4	2332	5	PCT-US99-11202-2	Sequence 2, Appl
40	285.5	9.4	2351	1	US-08-121-202-2	Sequence 2, Appl
41	285.5	9.4	2332	4	US-08-366-851A-2	Sequence 2, Appl
42	285.5	9.4	2332	4	US-09-324-867-3	Sequence 3, Appl
43	284.5	9.3	218	2	US-07-607-538C-5	Sequence 5, Appl
44	279.5	9.2	160	2	US-08-162-402B-15	Sequence 15, Appl
45	276.5	9.1	2304	4	US-09-324-867-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-08-936-135-18
Sequence 18, Application US/0893615
Patent No. 6054293

GENERAL INFORMATION:
 APPLICANT: He, Zhigang
 APPLICANT: Chen, Hang
 TITLE OF INVENTION: Semaphorin Receptors
 APPLICANT: Tessier-Lavigne, Marc
 NUMBER OF SEQUENCES: 26
 CORRESPONDENCE ADDRESS:
  ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
  STREET: 75 DENISE DRIVE
  CITY: HILLSBOROUGH
  STATE: CALIFORNIA
  COUNTRY: USA
  ZIP: 94010
 COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/936,135
  FILING DATE:
  CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
  NAME: OSMAN, RICHARD A
  REGISTRATION NUMBER: 36,627
  REFERENCE/DOCKET NUMBER: UC97-288-2
 TELECOMMUNICATION INFORMATION:
  TELEPHONE: (650) 343-4341
  TELEFAX: (650) 343-4342
 INFORMATION FOR SEQ ID NO: 18:
  SEQUENCE CHARACTERISTICS:
   LENGTH: 909 amino acids
   TYPE: amino acid
   STRANDEDNESS: single
   TOPOLOGY: linear
  MOLECULE TYPE: peptide
US-08-936-135-18
```

```
Query Match        11.4%;  Score 349;  DB 3;  Length 909;
Best Local Similarity   25.3%;  Pred. No. 8.6e-24;
Matches 109;  Conservative 72;  Mismatches 166;  Indels 84;  Gaps 16;

Qy   88 CGHLVTYQDSGTMTSKNYPGTVPNHTVCEKIITVPK-GKRRLIRLG-DLDIESQTCASDY 145
        ||:|::|:|||:|||:|||:|:||:|:|||:||:||| :|:||||:| :|| : ::||||:
Db   28 CGGRLNSKDAGTITSPGYPQDYPSHQNCEWIVYAPENQKIIVLNPNHFEIEKHDCKYDF 87
```

RESULT 2
US-08-936-135-20
Sequence 20, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
  APPLICANT: Tessier-Lavigne, Marc
  APPLICANT: He, Zhigang
  APPLICANT: Chen, Hang
  TITLE OF INVENTION: Semaphorin Receptors
  NUMBER OF SEQUENCES: 26
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
    STREET: 75 DENISE DRIVE
    CITY: HILLSBOROUGH
    STATE: CALIFORNIA
    COUNTRY: USA
    ZIP: 94010
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/936,135
    FILING DATE:
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: OSMAN, RICHARD A
    REGISTRATION NUMBER: 36,627
    REFERENCE/DOCKET NUMBER: UC97-288-2
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (650) 343-4341
    TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 20:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 926 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: peptide

```
Qy  146  LLF----TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY----AS  196
              ::::|   ::  :|                        :  |:||
Db   88  IEIRDGDSESADLLGKHGCNI-APPTIISSGSMLYIKFTSDYARQGAGFSLRYEIFKTGS  146

Qy  197  SD------------HPDLITC-----LERASHYLKTSYSKF------------------  220
                          |::|:      |:: :|::| :|
Db  147  EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPKMEII-LQFLIFDLEHDPLQVGEG  206

Qy  221  -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRYEG  277
            ||.:|   ||   |: ||.: |.:| ||  |:|| |:: |.:|  :|:|.|.|:|.||
Db  207  DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSELRSSTGILSLTFHTDMAVAK-----DC  259

Qy  278  ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE-----PDGQIRASSWQSVNESGDQVHHS  333
                                                   .:|:.||||  :||:|.|.:||
Db  260  CNVPLGMESGRIANEQISASSTY-----------------------------SDGRWT  305

Qy  334  PGQARLQDQGPSNASGDSSNNHKPREWLEIDLGEKKKITGIRTTG--STQSNFNFYVKSF  391
                                                                       |
Db  306  PQQSRLHGDDNGWTPNLDSN----KEYLQVDLRFLTMLTAIATQGAISRETQNGIYVKSY  361

Qy  392  VMNFKNNNSKKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIA  451
         |:| |.| :|||:|||:.:::.:|.||.::        ::|:|.|::|::|::|.|.||
Db  362  KLEVSTNGEDHMVYRHGKNI--RVFQANNDATEVVLNKLHAPLLTRFVRIRPQTWHSGIA  419

Qy  452  LKVELIGCQIT  462
         | :||.||::|
Db  420  LRLELFGCRVT  430
```

RESULT 3
US-08-936-135-22
Sequence 22, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
  APPLICANT: Tessier-Lavigne, Marc
  APPLICANT: He, Zhigang
  APPLICANT: Chen, Hang
  TITLE OF INVENTION: Semaphorin Receptors
  NUMBER OF SEQUENCES: 26
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
    STREET: 75 DENISE DRIVE
    CITY: HILLSBOROUGH
    STATE: CALIFORNIA
    COUNTRY: USA
    ZIP: 94010
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/936,135
    FILING DATE:
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: OSMAN, RICHARD A
    REGISTRATION NUMBER: 36,627
    REFERENCE/DOCKET NUMBER: UC97-288-2
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (650) 343-4341
    TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 22:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 901 amino acids

Query Match   11.4%;   Score 349;   DB 3;   Length 926;
Best Local Similarity   25.3%;   Pred. No. 8.9e-24;
Matches 109;   Conservative 72;   Mismatches 166;   Indels 84;   Gaps 16;

```
Db   88  CGHLVTYQDSGGTMTSKNYPGTYNHTVCEKTITVPK-GKRLLRLG-DLDIESQTCASDY  145
Qy   28  CGGRLNSKDAGYITSPGYPQDIYPSHQNCMIVAPEDMDKIVLNFNPHPFEIKHDCKYDF   87

Db  146  LLF----TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY----AS  196
Qy   88  IEIRDGDSESADLLGKHGCNI-APPTIISSGSMLYIKFTSDYARQGAGFSLRYEIFKTGS  146

Db  197  SD------------HPDLITC-----LERASHYLKTSYSKF------------------  220
Qy  147  EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPKMEII-LQFLIFDLEHDPLQVGEG  206

Db  221  -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRYEG  277
Qy  207  DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSELRSSTGILSLTFHTDMAVAK-----DG  259

Db  278  ILANGVLSRDGSLSDKRFLFTSNGCSRSLSFE-----PDGQIRASSWQSVNESGDQVHHS  333
Qy  260  CNVPLGMESGRIANEQISASSTY-----------------------------SDGRWT  305

Db  334  PGQARLQDQGPSNASGDSSNNHKPREWLEIDLGEKKKITGIRTTG--STQSNFNFYVKSF  391
Qy  306  PQQSRLHGDDNGWTPNLDSN----KEYLQVDLRFLTMLTAIATQGAISRETQNGIYVKSY  361

Db  392  VMNFKNNNSKKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIVARYVRVVPQTWHQRIA  451
Qy  362  KLEVSTNGEDHMVYRHGKNI--RVFQANNDATEVVLNKLHAPLLTRFVRIRPQTWHSGIA  419

Db  452  LKVELIGCQIT  462
Qy  420  LRLELFGCRVT  430
```

Left column:

```
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-936-135-22

Query Match   Best Local Similarity   23.2%;   Score 343.5;   DB 3;   Length 901;
Matches   111;   Conservative   72;   Pred. No. 2.8e-23;   Mismatches 177;   Indels 15;   Gaps
```

```
Qy    88  CGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
              |:||::||:||:|:|::||:|:||| :||:|   :|   |:|:|:||
Db    28  CGGRPNSKDAGYITSPGYPQDYFSHQNCEWIVYAPEPNQKIVLNFNPHFEIEKHDCKYDF  87

Qy   146  LLF---TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLITY-------AS 196
              ||:   |:||:|||||:|:||:|:||:|::|:|||::|:||::|:|::|:|    |:
Db    88  IEIRDGDSESADLLGKHCGNI-APPTIISSGSVLIYIKFTSDYARQGAGFSLRYEIFKTGS 146

Qy   197  SD------------------------------HPDLITC-----LERASHYLKTEYSKF--- 220
                                          :|||||      |:|::|:::||:::
Db   147  EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTIIAKPRMEIILQFLTFDLEHDPLQVGEG 206

Qy   221  -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGGISVLQRKGISRY-- 275
                ||:|:|:|:|   ::|:::|:|:||:::||:::||:|::||::|:::|::|:|:||
Db   207  DCKYDWLDIWDGIPHVGPLIGKYCGTKIPSKLRSSTGISLSLTFPHTDMAVAKDGFSARYYL 266

Qy   276  EGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGQIRAS 317
              :|:||||:|::|:|:::|::||:|:||:|::|::|||::||:|
Db   267  IHQEPPENFQCNVPLGMESGRIANEQISASSTFSDGR 303

Qy   318  SSWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSNHHKPREWLEIDLGEKKKTITGIRTT 377
              ||:|||:||:::|:||:|||||:|::|::|:||:||||:|::||:|:|:||:||:|:||:
Db   304  WTPQQSRLHGDDNGWTPNLDSN----KEYLQVDLRFLTMLTAIATQ 345

Qy   378  G--STQSNFNFYVKSFVWNFKNNNSKRKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIV 435
              |  :|::||:||:|:|:|::||::|:|:||:||:|||::|::|:|::||:|:||:|||||
Db   346  GAISRETQKGTYVKSYKLEVSTNGEDMMYTRHGKNH--KIFQANNDATEVVLNKLHMPLL 403

Qy   436  ARYVRVVPQTWHQRIALKVELIGCQITQGNDSLVWRKTS---QSTSVSTKKEDETITRP 491
              ||||||:||:||::|||::|:|:|::|::|||:||::||    :||:|:|||:|:||
Db   404  TRFIRIRPQTWHLGIALRLELFGCRVTDAPCSNMLGMLSGLIADTQISASSTREYLMSP 462
```

Right column:

```
REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (650) 343-4341
   TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 24:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 906 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: peptide
US-08-936-135-24

Query Match   Best Local Similarity   23.2%;   Score 343.5;   DB 3;   Length 906;
Matches   111;   Conservative   72;   Pred. No. 2.8e-23;   Mismatches 177;   Indels 119;   Gaps 15;
```

```
Qy    88  CGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
Db    28  CGGRPNSKDAGYITSPGYPQDYFSHQNCEWIVYAPEPNQKIVLNFNPHFEIEKHDCKYDF  87
Qy   146  LLF---TSSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLITY-------AS 196
Db    88  IEIRDGDSESADLLGKHCGNI-APPTIISSGSVLIYIKFTSDYARQGAGFSLRYEIFKTGS 146
Qy   197  SD------------------------------HPDLITC-----LERASHYLKTEYSKF 220
Db   147  EDCSKNFTSPNGTIESPGFPEKYPHNLDCTFTIIAKPRMEIILQFLTFDLEHDPLQVGEG 206
Qy   221  -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGGISVLQRKGISRY-- 275
Db   207  DCKYDWLDIWDGIPHVGPLIGKYCGTKIPSKLRSSTGISLSLTFPHTDMAVAKDGFSARYYL 266
Qy   276  EGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGQIRAS 317
Db   267  IHQEPPENFQCNVPLGMESGRIANEQISASSTFSDGR 303
Qy   318  SSWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSNHHKPREWLEIDLGEKKKTITGIRTT 377
Db   304  WTPQQSRLHGDDNGWTPNLDSN----KEYLQVDLRFLTMLTAIATQ 345
Qy   378  G--STQSNFNFYVKSFVWNFKNNNSKRKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIV 435
Db   346  GAISRETQKGTYVKSYKLEVSTNGEDMMYTRHGKNH--KIFQANNDATEVVLNKLHMPLL 403
Qy   436  ARYVRVVPQTWHQRIALKVELIGCQITQGNDSLVWRKTS---QSTSVSTKKEDETITRP 491
Db   404  TRFIRIRPQTWHLGIALRLELFGCRVTDAPCSNMLGMLSGLIADTQISASSTREYLMSP 462
```

RESULT 4
US-08-936-135-24
Sequence 24, Application US/0893613S
Patent No. 6054293
GENERAL INFORMATION:
   APPLICANT: Tessier-Lavigne, Marc
   APPLICANT: He, Zhigang
   APPLICANT: Chen, Hang
   TITLE OF INVENTION: Semaphorin Receptors
   NUMBER OF SEQUENCES: 26
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
      STREET: 75 DENISE DRIVE
      CITY: HILLSBOROUGH
      STATE: CALIFORNIA
      COUNTRY: USA
      ZIP: 94010
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30

RESULT 5
US-08-936-135-8
Sequence 8, Application US/0893613S
Patent No. 6054293
GENERAL INFORMATION:
   APPLICANT: Tessier-Lavigne, Marc
   APPLICANT: He, Zhigang
   APPLICANT: Chen, Hang
   TITLE OF INVENTION: Semaphorin Receptors
   NUMBER OF SEQUENCES: 26
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
      STREET: 75 DENISE DRIVE
      CITY: HILLSBOROUGH
      STATE: CALIFORNIA
      COUNTRY: USA
      ZIP: 94010
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30

Bottom of left column:

RESULT 4
US-08-936-135-24
Sequence 24, Application US/0893613S
Patent No. 6054293
GENERAL INFORMATION:
   APPLICANT: Tessier-Lavigne, Marc
   APPLICANT: He, Zhigang
   APPLICANT: Chen, Hang
   TITLE OF INVENTION: Semaphorin Receptors
   NUMBER OF SEQUENCES: 26
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
      STREET: 75 DENISE DRIVE
      CITY: HILLSBOROUGH
      STATE: CALIFORNIA
      COUNTRY: USA
      ZIP: 94010
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/936,135
      FILING DATE:
   ATTORNEY/AGENT INFORMATION:
      NAME: OSMAN, RICHARD A
      REGISTRATION NUMBER: 36,627
      CLASSIFICATION: 435

COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: Patentin Release #1.0, Version #1.30

RESULT 6

GENERAL INFORMATION:
  APPLICANT: Tessier-Lavigne, Marc
  APPLICANT: Chen, Hang
TITLE OF INVENTION: Semaphorin Receptors
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
  ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
  STREET: 75 DENISE DRIVE
  CITY: HILLSBOROUGH
  STATE: CALIFORNIA
US-08-936-135-10
Sequence 10, Application US/08936135
Patent No. 6054293

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/936,135
  FILING DATE:
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: OSMAN, RICHARD A.
  REGISTRATION NUMBER: 36,627
  REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (650) 343-4341
  TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 8:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 909 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: peptide
US-08-936-135-8

Query Match
Best Local Similarity   11.3%;   Score 343.5; DB 3; Length 909;
Matches   111; Conservative   72; Mismatches 177; Indels 119; Gaps 15;

```
Qy  88 CGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
          ::|:|  ||   |  :| |::|::|:  |::  |  |:||||:|| :|:||::| ||:
Db  28 CGGRPNSKDAGYIITSPGYPQDYPSHQNCEWIVYAPEPNQKIVLNFNPHFEIEKHDCKYDE 87

Qy 146 LLF------TSSSDQYGPYCGSMTVPKELLLNTSEVTVRPESGSHISGRGFLLTY---AS 196
Db  88 IEIRDGDSESADLLGKHGNI--APPTIISSGSVLYIKFTSDYARQGAGFSLRYEIFKTGS 146

Qy 197 SD-------HPDLITC-----LERASHYLKTEYSKF--------- 220
Db 147 EDCSKNFTSPNCTIESPGFPEKYPHINLDCTFTILAKPRMEIILQFLTFDLEHDPLQVGEG 206

Qy 221 -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRY-- 275
Db 207 DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSKLRSSTGILSLTFHTDMAVAKDGFSARYL 266

Qy 276 EGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGQIRAS 317
Db 267 IHQEPPENFQCNVPLGMESGRIANEQISASSTFSDGR------303

Qy 318 SSWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTT 377
Db 304 -----------WTPQQSRLRGDDGNGWTPNLDSN---KEYLQVDLRFLTMLTAIATQ 345

Qy 378 G--STQSNFNEYVKSFPVMNFKNNNSKKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIV 435
Db 346 GAISRETQKGYIVKSYKLEVSTNGEDWMVYRHGKNI--KIFQANNDATEVVLNRLHMPLL 403

Qy 436 ARYVRVVPQTWHRQRIALKVELLGCQITQGNDSLVWRKTS---QSTSVSTKKEDETITRP 491
Db 404 TFTIRIRPQTWHLGIALRLELFGCRVTDAPCSNMLGMLSGLIADTQISASSTREYLWSP 462
```

RESULT 7

GENERAL INFORMATION:
  APPLICANT: Tessier-Lavigne, Marc
  APPLICANT: He, Zhigang
  APPLICANT: Chen, Hang
US-08-936-135-12
Sequence 12, Application US/08936135
Patent No. 6054293

COUNTRY: USA
  ZIP: 94010
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/936,135
  FILING DATE:
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: OSMAN, RICHARD A.
  REGISTRATION NUMBER: 36,627
  REFERENCE/DOCKET NUMBER: UC97-288-2
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (650) 343-4341
  TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 10:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 909 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: peptide
US-08-936-135-10

Query Match
Best Local Similarity   11.3%;   Score 343.5; DB 3; Length 909;
Matches   111; Conservative   72; Mismatches 177; Indels 119; Gaps 15;

```
Qy  88 CGHLVTYQDSGTMTSKNY PGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
          ::|:|  ||   |  :| |::|::|:  |::  |  |:||||:|| :|:||::| ||:
Db  28 CGGRPNSKDAGYIITSPGYPQDYPSHQNCEWIVYAPEPNQKIVLNFNPHFEIEKHDCKYDE 87

Qy 146 LLF------TSSSDQYGPYCGSMTVPKELLLNTSEVTVRPESGSHISGRGFLLTY---AS 196
Db  88 IEIRDGDSESADLLGKHGNI--APPTIISSGSVLYIKFTSDYARQGAGFSLRYEIFKTGS 146

Qy 197 SD-------HPDLITC-----LERASHYLKTEYSKF--------- 220
Db 147 EDCSKNFTSPNCTIESPGFPEKYPHINLDCTFTILAKPRMEIILQFLTFDLEHDPLQVGEG 206

Qy 221 -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRY-- 275
Db 207 DCKYDWLDIWDGIPHVGPLIGKYCGTKTPSKLRSSTGILSLTFHTDMAVAKDGFSARYL 266

Qy 276 EGILANGVLSRDGSLSDKRFLFTSNGCSRSLSFEPDGQIRAS 317
Db 267 IHQEPPENFQCNVPLGMESGRIANEQISASSTFSDGR------303

Qy 318 SSWQSVNESGDQVHWSPGQARLQDQGPSWASGDSSNNHKPREWLEIDLGEKKKITGIRTT 377
Db 304 -----------WTPQQSRLRGDDGNGWTPNLDSN---KEYLQVDLRFLTMLTAIATQ 345

Qy 378 G--STQSNFNEYVKSFPVMNFKNNNSKKTYKGIVNNEEKVFQGNSNFRDPVQNNFIPPIV 435
Db 346 GAISRETQKGYIVKSYKLEVSTNGEDWMVYRHGKNI--KIFQANNDATEVVLNRLHMPLL 403

Qy 436 ARYVRVVPQTWHRQRIALKVELLGCQITQGNDSLVWRKTS---QSTSVSTKKEDETITRP 491
Db 404 TFTIRIRPQTWHLGIALRLELFGCRVTDAPCSNMLGMLSGLIADTQISASSTREYLWSP 462
```

RESULT 8

~ TITLE OF INVENTION: Semaphorin Receptors
~ NUMBER OF SEQUENCES: 26
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
~ STREET: 75 DENISE DRIVE
~ CITY: HILLSBOROUGH
~ STATE: CALIFORNIA
~ COUNTRY: USA
~ ZIP: 94010
~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: PatentIn Release #1.0, Version #1.30
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/08/936,135
~ FILING DATE:
~ CLASSIFICATION: 435
~ ATTORNEY/AGENT INFORMATION:
~ NAME: OSMAN, RICHARD A
~ REGISTRATION NUMBER: 36,627
~ REFERENCE/DOCKET NUMBER: UC97-208-2
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (650) 343-4341
~ TELEFAX: (650) 343-4342
~ INFORMATION FOR SEQ ID NO: 12:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 914 amino acids
~ TYPE: amino acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ MOLECULE TYPE: peptide

US-08-936-135-12

Query Match   11.3%;   Score 343.5; DB 3;   Length 914;
Best Local Similarity 23.2%; Pred. No. 2.9e-23;
Matches 111; Conservative 72; Mismatches 177; Indels 119; Gaps 15;

```
Qy    88 CGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
          ||  ::  |  | :| :||     :| |   |: ||   |::  | ||  |  :| |::
Db    28 CGGRPNSRDAGYITSPGYFQDDYPSHQNCEMIVTAPEFNQKIVLNFNPHFEIEKHDCKYDF 87

Qy   146 LIF----TSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY-----AS 196
          | |    | ||  |  :|  | |   | |: |  |  |  : | |  | |   |     ||
Db    88 IEIRDGDSESADLLGKHCGNI-APPTIISSGSVLYIKFTSDYARQGAGFSLRYEIFKTGS 146

Qy   197 SD--------------------HPDLITC----LERASHYLKTEYSKF----------- 220
          ::|                       |:|    || ||     |: |
Db   147 EDCSKRNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPRMEIILQFLTFDLEHDPLQVGEG 206

Qy   221 -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRY-- 275
               | |  ::    :  :| |     :|  ||  :  |:| |  |  : :|  |:|
Db   207 DCKYDWLDIWDGIPHVGPLIGKYCGTKPSKLRSSTGILSLTFHTDMAVAKDGFSARYYL 266

Qy   276 ----------EGILANGVLSRDGSLSDKRFLFTSNCGSRSLSFEPDGQIRAS 317
                     :| | | |   | |  | |||  || | | | |
Db   267 IHQEPPENFQCNVPLGMESGRIANEQISASSTFSDGR 303

Qy   318 SSWQSVNESGDDVHHSPGQARLQDQGPSWASGDSSNHKPREMLEIDLGEKKKITGIRTT 377
          |:  | :| || | ||   :|| |     |  |  ||   | | :| | |  :| :|||
Db   304 -WTPQQSRLHGDDNGWTPNLDSN---KEYLQVDLRFLTMLTAIATQ 345

Qy   378 G--STQSNFNTYVKSFVMNFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQONFIPPIV 435
           |  | | :|   |  |  |  |   |  |  |  ||   || | ||  | | | |||:||
Db   346 GAISRETQKGYVVKSKLEVSTNGEDMMVYRHGKNH--KIFQANNDATEVVLNKLHMPLL 403

Qy   436 ARYVRVVPQTWHQRIALKVEHLGCOITQGNDSLVWRKTS---QSTSVVSTKREDETITRP 491
          |: |||| | |  |  | |  :|  |  ||   :  | :    |  |: :   |:  |  |
Db   404 TRFIRIRPQTWHLGIALRLELFGCRVTDAPCSNMLGMLSGLIADPTQISASSTREYLWSP 462
```

US-09-691-344-14

~ Sequence 14, Application US/08936135
~ Patent No. 6054293
~ GENERAL INFORMATION:
~ APPLICANT: Tessier-Lavigne, Marc
~ APPLICANT: He, Zhigang
~ APPLICANT: Chen, Hang
~ TITLE OF INVENTION: Semaphorin Receptors
~ NUMBER OF SEQUENCES: 26
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
~ STREET: 75 DENISE DRIVE
~ CITY: HILLSBOROUGH
~ STATE: CALIFORNIA
~ COUNTRY: USA
~ ZIP: 94010
~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: PatentIn Release #1.0, Version #1.30
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/08/936,135
~ FILING DATE:
~ CLASSIFICATION: 435
~ ATTORNEY/AGENT INFORMATION:
~ NAME: OSMAN, RICHARD A
~ REGISTRATION NUMBER: 36,627
~ REFERENCE/DOCKET NUMBER: UC97-208-2
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: (650) 343-4341
~ TELEFAX: (650) 343-4342
~ INFORMATION FOR SEQ ID NO: 14:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 926 amino acids
~ TYPE: amino acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ MOLECULE TYPE: peptide

US-08-936-135-14

Query Match   11.3%;   Score 343.5; DB 3;   Length 926;
Best Local Similarity 23.2%; Pred. No. 2.9e-23;
Matches 111; Conservative 72; Mismatches 177; Indels 119; Gaps 15;

```
Qy    88 CGHLVTYQDSGTMTSKNYPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
          ||  ::  |  | :| :||     :| |   |: ||   |::  | ||  |  :| |::
Db    28 CGGRPNSRDAGYITSPGYFQDDYPSHQNCEMIVTAPEFNQKIVLNFNPHFEIEKHDCKYDF 87

Qy   146 LIF----TSSDQYGPYCGSMTVPKELLLNTSEVTVRFESGSHISGRGFLLTY-----AS 196
          | |    | ||  |  :|  | |   | |: |  |  |  : | |  | |   |     ||
Db    88 IEIRDGDSESADLLGKHCGNI-APPTIISSGSVLYIKFTSDYARQGAGFSLRYEIFKTGS 146

Qy   197 SD--------------------HPDLITC----LERASHYLKTEYSKF----------- 220
          ::|                       |:|    || ||     |: |
Db   147 EDCSKRNFTSPNGTIESPGFPEKYPHNLDCTFTILAKPRMEIILQFLTFDLEHDPLQVGEG 206

Qy   221 -CPAGCRDVAGDIS--GNMVDGYRDTSLLCKAAIHAGIIADELGGQISVLQRKGISRY-- 275
               | |  ::    :  :| |     :|  ||  :  |:| |  |  : :|  |:|
Db   207 DCKYDWLDIWDGIPHVGPLIGKYCGTKPSKLRSSTGILSLTFHTDMAVAKDGFSARYYL 266

Qy   276 ----------EGILANGVLSRDGSLSDKRFLFTSNCGSRSLSFEPDGQIRAS 317
                     :| | | |   | |  | |||  || | | | |
Db   267 IHQEPPENFQCNVPLGMESGRIANEQISASSTFSDGR 303

Qy   318 SSWQSVNESGDDVHHSPGQARLQDQGPSWASGDSSNHKPREMLEIDLGEKKKITGIRTT 377
          |:  | :| || | ||   :|| |     |  |  ||   | | :| | |  :| :|||
Db   304 -WTPQQSRLHGDDNGWTPNLDSN---KEYLQVDLRFLTMLTAIATQ 345

Qy   378 G--STQSNFNTYVKSFVMNFKNNNSKWKTYKGIVNNEEKVFQGNSNFRDPVQONFIPPIV 435
           |  | | :|   |  |  |  |   |  |  |  ||   || | ||  | | | |||:||
Db   346 GAISRETQKGYVVKSKLEVSTNGEDMMVYRHGKNH--KIFQANNDATEVVLNKLHMPLL 403
```

RESULT 9
US-08-936-135-16
Sequence 16, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
  APPLICANT: Tessier-Lavigne, Marc
  APPLICANT: Chen, Hang
  APPLICANT: He, Zhigang
  TITLE OF INVENTION: Semaphorin Receptors
  NUMBER OF SEQUENCES: 26
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
    STREET: 75 DENISE DRIVE
    CITY: HILLSBOROUGH
    STATE: CALIFORNIA
    COUNTRY: USA
    ZIP: 94010
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/936,135
    FILING DATE:
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: OSMAN, RICHARD A
    REGISTRATION NUMBER: 36,627
    REFERENCE/DOCKET NUMBER: UC97-288-2
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (650) 343-4341
    TELEFAX: (650) 343-4342
  INFORMATION FOR SEQ ID NO: 16:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 931 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: peptide

Query Match
Best Local Similarity   23.2%;   Score 343.5; Pred. No. 3e-23; DB 3; Length 931;
Matches 111; Conservative 72; Mismatches 177; Indels 119; Gaps 15;

Qy 436 ARVRVVPQTWHQRIALKVELIGCGIITQGNDSLVWRKTS---QSTSVSTKKEDETTIRP 491
         |||| ::| ||:||||::|||:|| :| ||| |:|:   :||:|:|||||| ||||
Db 404 TRFIRIRPQTWHLGIALRLELPGCRVTDAPCSNMLGMLSGLIADTQISASSTREYLMSP 462

Qy   88 CGHLVTIYQDSGTMTSKNFPGTYPNHTVCEKTITVPK-GKRLILRLG-DLDIESQTCASDY 145
         || : |: :|:|::|||:|| : |:::| : ||::| | | :| ::| ||:
Db   28 CGGRPNSRDAGYITSPGYPQDYFSHQNCEMIVYAPEPNQKIVLNPNPHPFEIEKHDCKYDF   87

Qy 146 LLF----TSSSDGYGPYCGSMTVPKELLNTSEVTVRFESGSSHISGRGFLLTY---AS 196
         : | ||||:| |||||||:||||| |::|| ::||:|| ||||||||| || ||
Db   88 IEIRGDGSESADLLGKHGGNI-APPILISSGSVLYIKFTSDYAROGAGFSLRYELFKTGS 146

Qy 197 SD-------------HPDLITC-----LERASHYLKTEYSKF------------ 220
                       |||||| |||||||||||||||||
Db 147 EDCSKNFTSPNGTIESPGFPEKPHNLDCTFTILAKPRMEIILQFLTFDLEHDPLQVGEG 206

Db 221 -CPAGCRDVAGCDIS--GNMVDGYRDYSLLCKAAIHAGIIADELGGGISVLQRKGISRY-- 275
Db 207 DCKYDMLDIWDGIFHVGPLIGKYCGTKTPSKLRSSTGILSLTFHTDMAVAKDGFSARYYL 266
Qy 267 IHQEPPENFQCNVPLGMESGRIANEQISASSTFSDGR---------- 303

RESULT 10
US-08-936-135-6
Sequence 6, Application US/08936135
Patent No. 6054293
GENERAL INFORMATION:
  APPLICANT: Tessier-Lavigne, Marc
  APPLICANT: He, Zhigang
  APPLICANT: Chen, Hang
  TITLE OF INVENTION: Semaphorin Receptors
  NUMBER OF SEQUENCES: 26
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
    STREET: 75 DENISE DRIVE
    CITY: HILLSBOROUGH
    STATE: CALIFORNIA
    COUNTRY: USA
    ZIP: 94010
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/936,135
    FILING DATE:
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: OSMAN, RICHARD A.
    REGISTRATION NUMBER: 36,627
    REFERENCE/DOCKET NUMBER: UC97-288-2
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (650) 343-4341
    TELEFAX: (650) 343-4342
  INFORMATION FOR SEQ ID NO: 6:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 923 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: peptide

Query Match
Best Local Similarity   24.0%;   Score 329; Pred. No. 6.8e-22; DB 3; Length 923;
Matches 113; Conservative 79; Mismatches 161; Indels 118; Gaps 21;

Qy 318 SSMQSVNESCDQVHRSPGSQARLQDQGPSWRASGDSSNNHKPREWLEIDLGEKKKTITGIRFT 377
         :::| | || :: :|:: | ||:||||||||:||:
Db 304 ------------WTPQQSRLHGDDNGWTPNLLDSN----KEYLQVDLRFLTMLTAIATQ 345

Qy 378 G--STQSNFNYFYKSFWHNRFKNNNSKWKTYIKGIVNNEEKYFQGNSNFRDPQVONNFIPPIY 435
         | | || | || ::| :: | |: | || |: ::|:: |:| :| ::|: :|
Db 346 GAISRETFQKGTYVKSYKLEVSTNGEDWRWYRHGKNH--KIFQANNDATEVVLNKLHMPLL 403

Qy 436 ARYRVRVVPQTWHQRIALKVELIGCGIITQGNDSLVWRKTS---QSTSVSTKKEDETTIRP 491
          | :| ||||||| ||| |||:||| :| | :|: | :||:|:||||| |||||
Db 404 TRFIRIRPQTWHLGIALRLELPGCRVTDAPCSNMLGMLSGLIADTQISASSTREYLMSP 462

Qy   64 RGL-------LALLIAVSAPLRLQAEELGDGGIHLVTYQDSGTMTSKNFPGTYPNHTVCEK 117
         ||| ||| || : | | :|| | :|::|: :|:|::|||:||:: |:||
Db   3 RGLPLLCATLALALALAGAFR-----SDKCGGTIKIENPGYLITSPGYPHSYHPSEKCEW   56

Qy 118 TITVPKG-KRLILRLG-DLDIESQTCASDYLFLTSSSDQ----YGPYCGSMTVPKELLIN 171
         :| || ::||:| | | :| ::| ||: | |||:| |||||||:|||||||:|
Db   57 LIQAPEPYQRIIINFNPHFDLEDRDCKYDVEVIDGENEGGRLWGKFCGKI-APSPVSS   115

Qy 172 TSEVTVRFESGSHISGRGFLLTYA--------SSDHPDLITC-- 205
         :|:|| ||:|| ||||||| ||| |||||||||
Db 116 GPFLFKFVSDYETHGAGFSIRYEIFRGPECSQNYAPTGVIKSPGFPEKYPNCLECTY 175

Qy 206 --------------LERASHYLKTEYSKFPAG---CRUVACDISGNMVDGTYRDYSLLCKAAIH 252

RESULT 11

US-07-607-538C-4

Sequence 4, Application US/07607538C

Patent No. 5455031

GENERAL INFORMATION:

  APPLICANT: Ceriani Dr., Roberto L.
  APPLICANT: Peterson Dr., Jerry A.
  APPLICANT: Larocca, David J.
  TITLE OF INVENTION: POLYPEPTIDE WITH 46
  TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFITY AND CLOTTING
  TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
  TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
  TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
  TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
  TITLE OF INVENTION: USE THEREOF
  NUMBER OF SEQUENCES: 5
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: V. Amzel & Assoc.
  STREET: 2055 No. 54550131th Broadway
  CITY: Walnut Creek
  STATE: California
  COUNTRY: USA
  ZIP: 94596
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
  SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/607,538C
  FILING DATE: 01-NOV-1990
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
  NAME: Viviana Amzel
  REGISTRATION NUMBER: 30,930
  REFERENCE/DOCKET NUMBER: CRPCC-004
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (510) 943-1931
  TELEFAX: (510) 943-1189
  TELEX: N.A.
  INFORMATION FOR SEQ ID NO: 4:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 218 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: protein
  FRAGMENT TYPE: linear

US-07-607-538C-4

Query Match   10.5%;   Score 321;   DB 1;   Length 218;
Best Local Similarity   42.6%;   Pred. No. 3.5e-22;

Qy 412 EEKVFQGNSNFRDPVQNNFIPPIVARYVRVVPGTWHQRIALKVELIGCOIT 462
Db 377 KAIIFQGNTNFPFDVVLGVFSKPLITREVFRIKPVSWEIGISMRFEVVGCKIT 427
Qy 323 ----KEHIQVDLGLLRVIYAVGTOGAISKETHKKYIVKTIYAVDLISSNGEDWISLK--EGN 376
Db 354 NHKPREWLEIDLGEKKKITGIRITG--STOSNFNFVKSFYMNFKNNNSKWKTYKGIVNN 411
Qy 275 ----CMEALGMESGEIHSDGDITASSQYGT-----NWSVERSRLMYPENGWPGEDSY 322
Db 298 ----TSNGCSRSLSFEP----DGQIRASSWQSVNESGCDQVHWSPGQARLQDQGPSWASGDSSN 353
Qy 224 IGRYCGQKPTPGRIRSSSGVLSWVFYTDSAIAKEGFSANYSVLQSSI-SEDFK------ 274
Db 253 AG-IIADELGCQI------------SVLQRGISRVEGILANGVLSRDGSLSDKRFLF 297
Db 176 IIFAPKMSEIILEFPSFPDLEQDSNP--PGGMPFCRYDRLEI-----WDGFPE-----VCPH 223

RESULT 12

US-08-162-402B-4

Sequence 4, Application US/08162402B

Patent No. 5972337

GENERAL INFORMATION:

  APPLICANT: CERIANI, ROBERTO L.
  APPLICANT: PETERSON, JERRY A.
  APPLICANT: LAROCCA, DAVID J.
  TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
  TITLE OF INVENTION: 46 KIDALTON HUMAN MILK FAT
  NUMBER OF SEQUENCES: 29
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Pretty, Schroeder & Poplawski
  STREET: 444 South Flower St., 19th Floor
  CITY: Los Angeles
  STATE: CA
  COUNTRY: USA
  ZIP: 90071
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: DOS
  SOFTWARE: FastSEQ for Windows Version 2.0
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/162,402B
  FILING DATE: 03-DEC-1993
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER:
  FILING DATE:
  ATTORNEY/AGENT INFORMATION:
  NAME: Amzel, Viviana
  REGISTRATION NUMBER: 30,930
  REFERENCE/DOCKET NUMBER: P66 38215
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 213-622-7700
  TELEFAX: 213-489-4210
  TELEX:
  INFORMATION FOR SEQ ID NO: 4:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 218 amino acids
  TYPE: amino acid
  STRANDEDNESS: unknown
  TOPOLOGY: unknown
  MOLECULE TYPE: peptide

US-08-162-402B-4

Query Match   10.5%;   Score 321;   DB 2;   Length 218;
Best Local Similarity   42.6%;   Pred. No. 3.5e-22;
Matches 72;   Conservative 30;   Mismatches 51;   Indels 16;   Gaps 5;

Qy 353 NNHKPREWLEIDLGEKKKITGIRTTGTGSTOSNFNFYVKRSFVMNFKNNNSKWKTYKGIVNNE 412
Db 58 NGCSTPLGMENGKIENKOITASSFKKSWW------GD-YWEPFRARLNAQGRVNAWQAKA 110
Qy 300 NGCSRSLSFFE-----PDGQIRASS---SWQSVNESGCDQVHWSPGQARLQDQGPSWASGDSS 352
Db 58 NGCSTPLGMENGKIENKOITASSFKKSWW------GD--YWEPFRARLNAQGRVNAWQAKA 110

RESULT 12

US-08-162-402B-4

Query Match   10.5%;   Score 321;   DB 2;   Length 218;
Best Local Similarity   42.6%;   Pred. No. 3.5e-22;
Matches 72;   Conservative 30;   Mismatches 51;   Indels 16;   Gaps 5;

Qy 300 NGCSRSLSFFE-----PDGQIRASS---SWQSVNESGCDQVHWSPGQARLQDQGPSWASGDSS 352
Db 58 NGCSTPLGMENGKIENKOITASSFKKSWW------GD--YWEPFRARLNAQGRVNAWQAKA 110
Qy 413 EKVFQGNSNFRDPVQNNFIPPIVARYVRVVPGTWHQRIALKVELIGCOI 461
Db 169 DKIFEGNTNIKGHVKNFFNPPIISRFIRVIPKTWNQSIALRLELFGCDI 217

```
Db  111  NNNK--QWLEIDLLKIKIIAIITQGCKSLSSEMYVRKSYTIHYSEGGYVEWKPYRLKSSMV 168
Qy  413  EKVPQGNSNFRDPVQQNNFIPPIYARYVRVVPQTWHQRIALKVELLGCOI 461
         :|:|:||| | |||||||:|:|:|:||:|||:|| |||
Db  169  DKIFEGNTNTKGHVKNFFNPPIISRFIRVIPKTWNQSIALRLELFGCDI 217
```

RESULT 13
; US-08-746-111-5
; Sequence 5, Application US/08746111
; Patent No. 6066778
; GENERAL INFORMATION:
; APPLICANT: Ginsburg, David
; APPLICANT: Cui, Jisong
; TITLE OF INVENTION: Compositions And Methods For Screening
; TITLE OF INVENTION: Compounds For Anticoagulant Activity
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,111
; FILING DATE: 06-NOV-1996
; CLASSIFICATION:
; CORRESPONDENCE ADDRESS:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: UM-02536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-746-111-5

Query Match 10.4%; Score 318; DB 3; Length 2183;
Best Local Similarity 40.8%; Pred. No. 3.1e-20;
Matches 69; Conservative 31; Mismatches 53; Indels 16; Gaps 5;

```
Qy  300  NCCSRSLSFE------PDGGIRASS----SWGSVNESGDGVHWSPGGARLODGGPSWASGDSS 352
Db 2023  NCCSTPLGLGLGDGRIODKGITASSFFKSWW----GD--YWEPSLARLNAGGRVNAWQAKA 2075
Qy  353  NHHKPREMLEIDLGEKKKITGIRTTGCSTQSNFNFYVKSFVMNFKNNNSKWKYYKGIVNNE 412
Db 2076  NNNK--QWLOVDLLKIKKIVNTAIVTQGCKSLSSEMYVKSYSIGYSDGGVANKPYRQKSSMV 2133
Qy  413  EKVPQGNSNFRDPVQNFIPPIYARVRVRVVPQTWHQRIALKVELIGCOI 461
Db 2134  DKIFEGNSNTKGHWKMNFNPPIISRFIRIIPKTWNQSIALRLELFGCDI 2182
```

RESULT 14
; US-08-162-402B-14
; Sequence 14, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.

; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-14

Query Match 10.2%; Score 311; DB 2; Length 160;
Best Local Similarity 41.7%; Pred. No. 1.8e-21;
Matches 70; Conservative 30; Mismatches 52; Indels 16; Gaps 5;

```
Qy  301  GCSRSLSFE------PDGGIRASS----SWGSVNESGDGVHWSPGGARLODGGPSWASGDSN 353
         :|:| |||| |||||||| |:||||||| ||||||||||||||||| ||
Db    1  GCSTPLGMENGKIENKGITASSFFKSWW----GD--YWEPFRARLNAGGRVNAWQAKAN 53
Qy  354  NHKPREMLEIDLGEKKKITGIRTTGCSTQSNFNFYVKSFVMNFKNNNSKWKYYKGIVNNE 413
Db   54  NNK--QWLEIDLLKIKKIIAIITQGCKSLSSEMYVKSYTIHYSEGGYVEWKPYRLKSSMVD 111
Qy  414  KVFQGNSNFRDPVQQNNFIPPIYARYVRVVPQTWHQRIALKVELIGCOI 461
Db  112  KIFEGNTNTKGHVKNFFNPPIISRFIRVIPKTWNQSITLRLELFGCDI 159
```

RESULT 15
; US-09-324-867-2
; Sequence 2, Application US/09324867A
; Patent No. 6251632
; GENERAL INFORMATION:
; APPLICANT: Lillicrap, David
; APPLICANT: Cameron, Cherie
; APPLICANT: No. 6251632ley, Colleen
; APPLICANT: Horrocks, L. Suzanne Hoyle
; APPLICANT: Hough, Christine
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
; FILE REFERENCE: 1669.0010002/JAG/BJD
; REGISTRATION NUMBER: US/09/324,867A
; CURRENT APPLICATION NUMBER: 09/035,141
; CURRENT FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER:

```
;  EARLIER FILING DATE: 1998-03-059
;  EARLIER APPLICATION NUMBER: 60/039,953
;  EARLIER FILING DATE: 1997-03-06
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: PatentIn Ver. 2.0
;
;  SEQ ID NO 2
;  LENGTH: 2343
;  TYPE: PRT
;  ORGANISM: Canis familiaris
US-09-324-867-2


Query Match           10.0%;   Score 306.5; DB 4;   Length 2343;
Best Local Similarity  40.5%;   Pred. No. 4,2e-19;
Matches  68;  Conservative  26;  Mismatches  61;  Indels  13;  Gaps  4;

Qy   300 NGCSRSLSFE-----PDGGIRASSSWQSVNESGDQVIWSPGQARLQDQGFSWASGDSSNNH 355
             |  ||  :||                ||      :        ||    ||:||  ||
Db  2183 NSCSMPLGMESKAISDAQITASSYLSSMLAT------WSPSQARLHLQGRTNAWRPQANN- 2236

Qy   356 KPREWLEIDLGEKKKITGIRPTGSSTQSNFNFYVKSFVMNFKNNNSKWKKTYKGIVNNEEKV 415
             |:|||::|  :  |:||:|    ||: :|||   || :  :   |  : ||:|||  ||
Db  2237 -PKEWLQVDFRKTMKVTGITTQGVKSLLISMYVKEFLISSSQDGHHNWTLF--LQNGKVKV 2293

Qy   416 FQGNSNFRDPVQNNFIPIVARYVRVVPQTWHQRIALKVELIGCQITQ 463
             ||||  ||| |||||:||:|||||||::||::|:|| |
Db  2294 FQGNRDSSTPVRNRLEPPLVARYVRLHPQSWAHHIALRLEVLGCDTCQ 2341
```

PD - 2/18/00   FD - 10/18/00

us-09-691-344a-3.rng

File copy p. 6 & 7

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 23:36:58 ; Search time 296.45 seconds
(without alignments)
10198.976 Million cell updates/sec

Title: US-09-691-344A-3
Perfect score: 1761
Sequence: 1 atggattcgtcgtggggca.........caagtgatgtcggcagttaa 1761

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*

```
 1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
 2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
 3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
 4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
 5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
 6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
 7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
 8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
 9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001b.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1761	100.0	1761	22	AAS00614	Novel human protei
2	1686.8	95.8	3594	22	AAS00660	Human TANGO 229 cDN
3	1620	92.0	1620	22	AAS00615	Novel human protei
4	1522	86.4	1768	22	AAS00613	Novel human protei
5	539.2	30.6	1871	21	AAB967736	Polynucleotide iso
6	221.8	12.6	2428	22	AAH34855	Human colon cancer
7	210.4	12.5	407	20	AAV87469	EST clone BP481.
8	205.6	11.7	2046	21	AAS1872	Human Factor 8 Hom
9	127.2	7.2	252	22	AAH36736	Human colon cancer

RESULT 1

AAS00614

ID   AAS00614 standard; cDNA; 1761 BP.
AC   AAS00614;
DT   29-AUG-2001
DE   Novel human protein (NHP) DNA sequence #2.

Novel human protein (NHP; CUB domain; extracellular domain; gene therapy;
obesity; high blood pressure; connective tissue disorder; infertility;
NHP-mediated pathway; ss.

Homo sapiens.

Key	Location/Qualifiers
CDS	1..1761
	/tag= a
	/product= "Novel human protein #2"

ALIGNMENTS

CC   The sequence represents a polynucleotide which encodes a novel human
CC   protein (NHP) containing a CUB domain (an extracellular domain). CUB
CC   proteins have been associated with regulating development, modulating
CC   cellular processes and preventing infectious disease. NHP nucleotide
CC   sequences are useful for gene therapy of physiological disorders or
CC   diseases. NHP oligonucleotides are useful as hybridisation probes for
CC   screening libraries and assessing gene patterns. NHP nucleotide sequences
CC   are useful for detecting mutant or inappropriately expressed NHPs (for
CC   example those proteins associated with obesity, high blood pressure,
CC   connective tissue disorders and infertility) for the diagnosis of a
CC   disease. The polynucleotides may also be used in screening for drugs
CC   effective in the treatment of symptomatic or phenotypic manifestations of
CC   perturbing the normal function of NHP in the body. Nucleotide constructs
CC   encoding NHP products are used to genetically engineer host cells to
CC   express such products in vivo. These host cells allow for the
CC   identification of compounds that bind to NHP receptors or trigger
CC   NHP-mediated pathways.
XX
PS
XX
SQ   Sequence 1761 BP; 471 A; 402 C; 483 G; 405 T; 0 other;

Query Match   100.0%; Score 1761; DB 22; Length 1761;
Best Local Similarity   100.0%;
  Matches 1761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 atgggattcgtgcgggggactgcgcccgctccggcgccccgtccggtcccgcaagg 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 atgggattcgtgcgggggactgcgcccgctccggcgccccgtccggtcccgcaagg 60

Qy         61 gaggcggccccgcccgccgggcagctcgcggggatccggtcagggggaggccggcc 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 gaggcggccccgcccgccgggcagctcgcggggatccggtcagggggaggccggcc 120

Qy        121 aagctggcgcgcccaggcgggtcatgtgtcccggcccgccgggcggcgggcactgg 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 aagctggcgcgcccaggcgggtcatgtgtcccggcccgccgggcggcgggcactgg 180

Qy        181 gctgcgccgggcgggggcctccgtgcttgtcgggtcccgccgggctcgggcgcacg 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 gctgcgccgggcgggggcctccgtgcttgtcgggtcccgccgggctcgggcgcacg 240

Qy        241 gcgggaggaggagctcgggtgatgctgtggacaccactacggagtagtgcacaatg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 gcgggaggaggagctcgggtgatgctgtggacaccactacggagtagtgcacaatg 300

Qy        301 acatctaagaattacccggacctaccactcacacactgctttcgcaagacaattaca 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 acatctaagaattacccggacctaccactcacacactgctttcgcaagacaattaca 360

Qy        361 gtaccaaagggggaaagaactgattctgaggtcgtggttggatcgaatcccagacc 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 gtaccaaagggggaaagaactgattctgaggtcgtggttggatcgaatcccagacc 420

Qy        421 tgtgctcctgactatctcttctccacagctcttcagatcaatatgtccatactgtgga 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 tgtgctcctgactatctcttctccacagctcttcagatcaatatgtccatactgtgga 480

Qy        481 agtatgactgtcccaaagaaactctgtgaaacacaagtgaatgaacgtccgcttcag 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 agtatgactgtcccaaagaaactctgtgaaacacaagtgaatgaacgtccgcttcag 540

Qy        541 agtggatcccacaattcctgccgggtcttttgctgacctatcgcgagcgagcgaccatcca 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 agtggatcccacaattcctgccgggtcttttgctgacctatcgcgagcgagcgaccatcca 600
```

SEQ ID NO:3
Databank; N_geneseq-032802
Az NO; AAS00660

```
RESULT  2

AAS00660

ID   AAS00660 standard; cDNA; 3594 BP.
XX
AC   AAS00660;
XX
DT   07-SEP-2001  (first entry)
XX
DE   Human TANGO 229 cDNA.
XX
KW   Human; TANGO 229; T cell; heart; liver; pancreas; placenta; brain; lung;
KW   skeletal muscle; kidney; spleen; lymph node; peripheral blood leukocyte;
KW   bone marrow; thymus tissue; abnormal blood coagulation; asthma; cancer;
KW   anaphylaxis; hepatitis; multiple sclerosis; coronary artery disease; ss;
KW   malaria; atopic dermatitis; amyotrophic lateral sclerosis; meningitis;
KW   attention deficit disorder; Crohn's disease; gastroenteritis; goitre;
KW   hypoglycaemia; diabetes mellitus; endometriosis; pulmonary embolism;
KW   muscular dystrophy; immuno-competence; vertebrate; blood; serum.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             72..2219
FT                   /tag= a
FT                   /product= "Human TANGO 229"
FT   sig_peptide     72..173
FT                   /tag= b
FT   mat_peptide     174..2216
FT                   /tag= c
FT                   /product= "Mature human TANGO 229"
XX
PN   WO200129088-A1.
XX
PD   26-APR-2001.
XX
PF   23-JUN-2000; 2000WO-US17386.
XX
PR   19-OCT-1999; 99US-0420707.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   MacKay CR, Myers PS, Kirst SJ, Fraser CC, Leiby KR;
XX
DR   WPI; 2001-308477/32.
DR   P-PSDB; AAU00670.
XX
PT   New isolated nucleic acid molecule for diagnosis, prevention, and
PT   therapy of human and other animal disorder, or as modulating agent for
PT   regulating cellular processes.
XX
PS   Claim 1; Fig 1; 263pp; English.
XX
CC   The sequence represents a cDNA which encodes human TANGO 229 polypeptide.
CC   This protein and similar others exhibit the ability to affect growth,
CC   proliferation, survival, differentiation, activity, morphology, or
CC   movement/migration of, e.g. T cells and cells of the heart, liver,
CC   pancreas, brain, lung, skeletal muscle, kidney, spleen, lymph
CC   node, peripheral blood leukocyte, bone marrow or thymus tissue. They can
CC   be used as modulating agents for regulating cellular processes, thus, the
CC   proteins and their associated nucleic acids can be used to prognosticate,
CC   prevent, diagnose, or treat disorders associated with physiological
CC   processes. These disorders include abnormal blood coagulation, asthma,
CC   anaphylaxis, hepatitis, multiple sclerosis, cancer, coronary artery
CC   disease, malaria, atopic dermatitis, amyotrophic lateral sclerosis,
CC   meningitis, attention deficit disorder, Crohn's disease, gastroenteritis,
CC   goitre, hypoglycaemia, diabetes mellitus, endometriosis, pulmonary
CC   embolism and muscular dystrophy. Antibodies to disorders such as these
CC   can be made by providing a polypeptide of the invention to an
CC   immuno-competent vertebrate and harvesting blood or serum from the
CC   vertebrate.
XX
SQ   Sequence 3594 BP; 958 A; 810 C; 921 G; 905 T; 0 other;

Query Match             95.8%;  Score 1686.8;  DB 22;  Length 3594;
Best Local Similarity   99.9%;  Pred. NO. 0;
Matches 1688;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy           71 ggccccggggcagctgcgcagtcgggatccgtcggagcttgccaagctggcgc 130
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 ggccccggggcagctgcgcagtcgggatccgtcggagcttgccaagctggcgc 60

Qy          131 ccagcgcgggtcatgtgcgcgcgcgggcgggcggcgggcgggcgggc 190
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db           61 ccagcgcgggtcatgtgcgcgcgcgggcgggcggcgggcgggcgggc 120

Qy          191 ggggcccctggctgcttgctgctgtcccggctccgcccgcccgcccgcccg 250
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db          121 ggggcccctggctgcttgctgctgtcccggctccgcccgcccgcccgcccg 180

Qy          251 tgggtagctggctgtgacacccagtgacttacctcatcaggatgggtggggag 310
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          181 tgggtagctggctgtgacacccagtgacttacctcatcaggatgggtggggag 240

Qy          311 atatcccgggacctacccccaatcacacgttgcggaaaagacaattacagcaaagg 370
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          241 atatcccgggacctacccccaatcacacgttgcggaaaagacaattacagtaccaaagg 300

Qy          371 ggaaaagactgatcctgaggtgtgggagatttggatatcgaatcccagaacctgctctg 430
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          301 ggaaaagactgatcctgaggtgtgggagatttggatatcgaatcccagaacctgctctg 360

Qy          431 actatctctctcaccagctcttcaccactggtgtcatggtcggaagtatgactg 490
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          361 actatctctctcaccagctcttcaccactggtgtcatggtcggaagtatgactg 420

Qy          491 ttcccaagaactcttgttggaacacaagtgaagtccgcttgagagtggatccc 550
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          421 ttcccaagaactcttgttggaacacaagtgaagtccgcttgagagtggatccc 480

Qy          551 acattctggccgggttcttctgggaatatgtgagttcagaggataattaataa 610
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          481 acattctggccgggttcttctgggaatatgtgagttcagaggataatataacct 540

Qy          611 catgttggaacgagctagcaggagacagcattctgggaataggtataaga 670
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          541 catgttggaacgagctagcaggagacagcattctgggaactaggtaaataga 600

Qy          671 gttgtagacgtagcaggaagacattctgggaatatgtgagtatagaggataact 730
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          601 gttgtagacgtagcaggaagacattctgggaactaagtgagttataagactt 660

Qy          731 ctttattcgtgcaaagctcgccatccatcgacaggaagaataattcgtcaagactaggt 790
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          661 cttttattcgtgcaaaagtcgccatccatcgacaggaataattcgtcaactaggt 720

Qy          791 tcagttgcttcagcgacaaaggacagtcgatattgaagggatcctggccaagtcgttc 850
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          721 tcagttgcttcagcgacaaaggatcacgtcgatattgaagggatcctggccaagtcgttc 780

Qy          851 tttccaagggtagttgttccctgtcagacaacgcatttctgttacccaatggtcagca 910
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          781 tttccaagggtagttgttccctgtcagacaacgcatttctgttacccaatggtcagca 840

Qy          911 gatcctgagtttgaacctgacgacgcaaatcagagatctcctcatgcagtcgtca 970
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db 1621 gaggctcagaaaacagactgtgggaagacagcagattaaatatccccttgccagcgatca 1680
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1681 gctggagtttaccatcagctcgatcaatgagaagaggagaggagacacaaaagttagatcc 1740
Db 1681 gctggagtttaccatcagctctgatcaatgagaagaggagaggagacacaaaagttagatcc 1740
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1741 acaagtggatatggcaggttaa 1761
Db 1741 acaagtggatatggcaggttaa 1761
```

```
Db   841  gatcctgagtcttgaacctgacgggaccaagtcagagcttcttccccatggcatcggtcca  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   901  atgcggagtcgggaccaagtcagagcttcactggttctctgtgccaagtcttcagagaccaaggcc  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  catcatgggctcggggcacagtagcacaacacaacaacaacaacagaggtggcctcggatcg  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1021  atttggggaggagaaaaagtcaagtcaacagtagcacaacacaacaacacaagaggtggcctcggatcg  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  tcaacttctttatgttaagcagtttgtgtgactccaaaaacaataattctaagtggaaga  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1141  ccttataaaggaatctgaataatgaagaaaggtgttcacctccaacttccgg  1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  accagtgcaaacaaattcatccccccacgtaggcggtgttgtcccc  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1261  gtaatgattccatcgtctggtcgccaagacaaagtcaaagcacaccagtgttccaactaagaaag  1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  gtaatgattccatcgtctggtcgccaagacaaagtcaaagcacaccagtgttccaactaagaaag  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1381  aagatagagacaatacacagctaggatgggttaaaacaacggaatcacacggataaacatta  1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  caacgtggctcattccattcctgtgtctcctggttgttgctggaatggggatct  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1501  ttgcagacccttagaagaagaagaagaagaaggtcgtgatcagcgggaggtccaga  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  ttgcagacccttagaagaagaagaagaaggtccgtatggacaggaggcccaga  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1621  ccatcagctatgtaatggaaagataatccccttgccagacaagcagtcagctgaagtta  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  tggcagatta  1690
          |||||||
Qy  1691  ccatcagctatgtaatggaaagattagaatctcatcacaagcgata  1750
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  aaacagactgttggaagcagattaaatatcccttgccagacacataagctttgaagtta  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1631  aaacagactgttggaagcagattaaatatcccttgccagacatcaagtcagctgaagtta  1690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  ttgcagacccttagaagaagaagaagaaggtccgtatggacaggaggcccaga  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1571  ttgcagacccttagaagaagaagaagaagtctatggatcagcgggaggtccaga  1630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  caacgtggctcattccattcctgtgtctcctggttgttgctggaatgggggatct  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1511  caacgtggctcattccattcctgtgtctcctggttgttgctggaatgggggatct  1570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  aagatagagacaatacacagctaggatgggttaaaacaacggaatcacacggataaacatta  1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1451  aagatagagacaatacacagctaggatgggttaaaacaacacggaatcacacggataaacatta  1510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  gtaatgattccatcgtctggtcgccaagacaaagtcaaagcacaccagtgttccaactaagaaag  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||,
Qy  1391  gtaatgatttccatcgtctggcaagacaaagtcaaagcacaccagtgttccaactaagaaag  1450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  accagtgcaaacaaattcatccccccacgtaggcggtgttgtcccc  1330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1331  agacatggcaccagaggatacgcctgaagctggagctcattggttgccagatcacacaag  1390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  agacatggcaccagaggatacgcctgaagctggagctcattggttgccagatcacacaag  1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1271  accagtgcaaacaaattcatccccccacgtaggcggtgttgtcccc  1330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  cctataaaggaatctgaataatgaagaaaggtgttcacctccaacttcggg  1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1211  cctataaaggaatctgaataatgaagaaaggtgttcacctccaacttcggg  1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  tcaacttctttatgttaagcagtttgtgtgactccaaaaacaataattctaagtggaaga  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1151  tcaacttctttatgttaagcagtttgtgtgactccaaaaacaataattctaagtggaaga  1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  atttggggaggagaaaaagtcaagtcaacaacacaacagaggtggcctcggatcg  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1091  atttggggaggagaaaaagtcaagtcaacagtagcaacaacacaacaacaagaggtggcctcggatcg  1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  catcatgggctcggggcacagtagcacaacacaacaacaacaacagaggtggcctcggatcg  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1031  catcatgggctcggggcacagtagcacaacacaacaacacaagaggtggcctcggatcg  1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  atgcggagtcgggaccaagtcagagcttcactggttctctgtgccaagtcttcagagaccaaggcc  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   971  atgcggagtcgggaccaagtcagagcttcactggttctctgtgccaagtcttcagagaccaaggcc  1030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT  3
AAS00615
ID    AAS00615 standard; cDNA; 1620 BP.
XX
AC    AAS00615;
XX
DT    29-AUG-2001  (first entry)
XX
DE    Novel human protein (NHP) DNA sequence #3.
XX
KW    Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW    obesity; high blood pressure; connective tissue disorder; infertility;
KW    NHP-mediated pathway; ss.
,
```

```
Query Match            92.0%;  Score 1620; DB 22;  Length 1620;
Best Local Similarity  100.0%;   Pred. No. 0;
Matches 1620;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    142  atgtgccggcccgggctccgcggcggcggcgcactggcgccggcggcggcctcctg  201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1   atggtgccggcgcgcccgcggcggcgcactggcgccggcggcgggcctcctg  60

Qy    202  gcttctgctcggtctccggcccccgggctcaaggcgggagagctggtgtggtgagtggc  261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  gcttctgctcggtctccgcggcccccgggctcaaggcgggagagctggtgtggtcggc  120

Qy    262  tgtggaacctagtgacttatcagaagtagtgcacacatgacatctaagaaattcatccggg  321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  tgtggaacctagtgacttatcagaagtagtgcacacatgacatctaagaaattcatccggg  180

Qy    322  acctaccccaatcacacactgtttgcgaaaagacaaattacagcagtgacacaaagcagactg  381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  acctaccccaatcacacactgtttgcgaaaagacaaattacagcagtgacacaaagcagactg  240

Qy    382  atttcgaggttcgggagattcgatcatcagaatcccgacctgtgctcctggactatcctcctc  441
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   241 attctgcagaggagaacattggatatcgaatccacgaccgtgctactgttctgactatctc 300

Qy   442 ttcacccagctcctcagatcacaatatgtgtcccacaaagaa                   501
         ||||||||||||||||||||||||||||||||||||||||||
Db   301 tccacccagctcctcagatcacaatatgtccatctggtactgttctgactatctggaa   360

Qy   502 cggcgccaaaggatcagtcgatatgaaggagatctggcacatccagttcctctcctggga 561
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ctcttgttgaaccacaagtgaagtaaccgtccgcttgagtggtggaacacttctggc     420

Qy   562 cggcgtttttgtttgatgaactcaacgatatgaaggagatctggcacatcagttatggga 621
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 cggggtttttttgtcttgactatgcagacagtgctttattaacacgttctggaa        480

Qy   622 csgctagccattatttgaacagacagaatacagacaaattctgccagcgtggttgtagagac 681
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 cgagctagccattattgaagacaagaacattctgccacattaacaccagttgaagaac    540

Qy   682 gtacgcagaggaacattcttggaatatgttggatatagagatatcttattattgtc     741
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 gtagcacgagacatttctgggatatatggtagatggaataacagatacctctattgtc   600

Qy   742 aaagctgccatccatcccatcatgccatgcttctgtgatgaatagtcaatcagtcgtgctt 801
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 aaagctgccatccatccggaataatgctgtcgatgtactagtggcttatgtgactgttcccaaagaa 660

Qy   802 cagcgcaaaggatcagtcgatatgaaggagatctggcacatcagttcctccaggggat   861
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 caccgcaaaggatcagtgatatgaaggagatctggcacaatggctgttctttcaaggggat 720

Qy   862 cagccacaggccgattctcgttcacctccaatggttgtgctgcagatcctcgagt      921
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ggttccctgccaacaggccgattctcgttcacctccaatggttgtgctgcagatcctcgagt 780

Qy   922 tttgaacctgcagcaaatcagagctcctcctcatgctggatcgatcgtcaatgagagtgga 981
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 tttgaacctgacgggcaaatcagagctctttcttccttcatatgctggatcgatcgtcaatgagagtgga 840

Qy   982 gaccaagttcactggtctcctggccgactcaaggccatctaggggct             1041
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 gaccaagtcactggtctcctgccgccgactcaaggccatctgggct             900

Qy  1042 tcgggcgacagtagcacacacaacaaaccacagagtcgtcgaagatcgattcgggaag    1101
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 tcgggcgacagtagcacacacaacaaaccacagagtcgtgagaagtcgatttggggaag   960

Qy  1102 aaaaaagaaataacagaagtaacacagtacacagctcttcaactctcaactttat     1161
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 aaaaaagaaaataacaagaatagaccacagtcgaactcacacagtcaa            1020

Qy  1162 gttaaagagtttgtgttgaactcagatgctcttcaaaaacaataattctaagtgaagga   1221
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 gttaagagttttgatgaactcaaagaacacaatattccctcgtgccagagacctatataagga 1080

Qy  1222 attgtgttaataatgaagaagaagtgttccagttacctaactaaagga             1281
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 cagaggatagtcctgaaggtggagctcattcgtccagtgattgcgccagtgcaa        1140

Qy  1282 aacaattcatcctccccaccttcgctggcacagaatacgctggttgtcctcccagaaccggcac 1341
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 aacaattcatcccctcccacccatcggcacagaattctggcac                 1200

Qy  1342 cagagagtagcctgaaggtggagctcattggtctgccagattacacaagtagtgatgattca 1401
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 cagaggagtagcctcgaaggctcattggttggtctgccagattacacaagtaatggatcca 1260

Qy  1402 ttggtgtgtggcgcagaacagtcaaagccacagtgtcttcaactcaagaagatgagaca    1461
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 ttggtgtggcgcagaacagtcaggtaactcaactcttgggaccaagtgcaa           1320

Qy  1462 atacaagagcccatccctcggaagacaaacattacaacgtggct                1521
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 atcacaaggcccatcccctcggaagaaacatccacacaggaatacaacattacaacgtggct 1380
```

RESULT 4

AAS00613

ID   AAS00613 standard; DNA; 1768 BP.
XX
AC   AAS00613;
XX
DT   29-AUG-2001 (first entry)
XX
DE   Novel human protein (NHP) DNA sequence #1.
XX
KW   Novel human protein; NHP; CUB domain; extracellular domain; gene therapy;
KW   obesity; high blood pressure; connective tissue disorder; infertility;
KW   NHP-mediated pathway; ds.
XX
OS   Homo sapiens.
XX
XX
FH   Key   Location/Qualifiers
FT   CDS   59..1522
FT   /tag= a
FT   /product= "Novel human protein #1"
XX

WO200129219-A1.
XX
PN   WO200129219-A1.
XX
PD   26-APR-2001.
XX
AP   08-OCT-2000; 2000WO-US28798.
XX
PR   19-OCT-1999; 99US-0160285.
PR   18-FEB-2000; 2000US-0183583.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Donoho G, Turner CA, Nehls M, Friedrich G, Zambrowicz B, Sands AT;
XX
PT   Novel nucleic acid encoding human protein, useful for drug screening, diagnosis and treatment of physiological disorders or diseases -
XX
PS   Disclosure: Page 30-31; 33pp; English.
XX
CC   WPI: 2001-290917/30.
CC   P-PSDB; AAU00628.
XX
CC   The sequence represents a polynucleotide which encodes a novel human
CC   protein (NHP), containing a CUB domain (an extracellular domain). CUB
CC   proteins have been associated with regulating development, modulating
CC   cellular processes and preventing infectious disease. NHP nucleotide
CC   sequences are useful for gene therapy of physiological disorders or
CC   diseases. NHP oligonucleotides are useful as hybridisation probes for
CC   screening libraries and assessing gene patterns. NHP nucleotide sequences
CC   are useful for detecting mutant or inappropriately expressed NHPs (for
CC   example, those proteins associated with obesity, high blood pressure,
CC   connective tissue disorders and infertility). NHP nucleotide probes (for
CC   the diagnosis of a disease associated with obesity, high blood pressure,
CC   connective tissue disorders and infertility). NHP nucleotide probes may
CC   also be used in screening for drugs
CC   effective in the treatment of symptomatic or phenotypic manifestations of

SEQ ID NO:3
AC NO: AAA96736, Database; N_Geneseq_032802

CC perturbing the normal function of NHP in the body. Nucleotide constructs
CC encoding such NHP products are used to genetically engineer host cells to
CC express such products in vivo. These host cells allow for the
CC identification of compounds that bind to NHP receptors or trigger
CC NHP-mediated pathways.
XX
SQ Sequence 1768 BP; 539 A; 370 C; 421 G; 438 T; 0 other;

Query Match   86.4%; Score 1522; DB 22; Length 1768;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   240 gcggagggaggagctggtgatgatggctggaccaccaggcacat 299
            ||||||||||||||||||||||||||||||||||||||||||||
Db   1 gcggagggaggagctggtgatgatggctggaccaccaggcacat 60

Qy   300 gacatcctaagaatttatcccgggaccatcacacactgttcgcaaagaaactac 359
Db   61 gacatcctaagaatttatcccgggaccatcacacactgttcgcaaagaaactac 120

Qy   360 agtacacaaggaggaaaagactgattctgagtttggagatatcgatcccagagc 419
Db   121 agtacacaaggaggaaaagactgattctgagtttggagatatcgatcccagagc 180

Qy   420 ctgtgttctgactactctcctccaccacgatcaatattggtctactactgtgg 479
Db   181 ctgtgttctgactactctcctccaccacgatcaatattggtctactactgtgg 240

Qy   480 aagtatgacgtgtccccaaagacaactcttgttgaacacaagtcgaactgaac 539
Db   241 aagtatgacgtgtccccaaagacaactcttgttgaacacaagtcgaactgaac 300

Qy   540 gaatggatcccacacttctcgcgcgggggttttttgctgacctatgcggagcga 599
Db   301 gaatggatcccacacttctcgcgcgggggttttttgctgacctatgcggagcga 360

Qy   600 agattttattaataacatgttggaacgacgctagccatattattgaagacagaga 659
Db   361 agattttattaataacatgttggaacgacgctagccatattattgaagacagaga 420

Qy   660 ctgccccagctggttgtagagacgtagcaggaggacatcttcgggaatcatgtgg 719
Db   421 ctgccccagctggttgtagagacgtagcaggaggacatcttcgggaatcatgtgg 480

Qy   720 tagagataccttctttattgtgcaaagccatcccatacagcagggaatcgtgaact 779
Db   481 tagagataccttctttattgtgcaaagccatcccatacagcagggaatcgtgaact 540

Qy   780 aaggtggccagatcaagtcgcttcagcgcaaaggactcagtcagtccaatgaaggg 839
Db   541 aaggtggccagatcaagtcgcttcagcgcaaaggactcagtcagtccaatgaaggg 600

Qy   840 caatggtgttcttcgaggatggttgcctctgtcagacaacgacga ttctgttca 899
Db   601 caatggtgttcttcgaggatggttgcctctgtcagacaacgacga ttctgttca 660

Qy   900 tggttgcgcagacagcaatccgagttctcctg 959
Db   661 tggttgcgcagacagcaatccgagttctcctg 720

Qy   960 gcagtcgtcaatgagagtgggaccaaggccatgg 1019
Db   721 gcagtcgtcaatgagagtgggaccaaggccatgg 780

Qy   1020 ggacaaagccccatccatgggcttcgggcgacagtgcaacagt 1079
Db   781 ggacaaagccccatccatgggcttcgggcgacagtgcaacagt 840

Qy   1080 gctggagatcgattttggggaaagaaaataacaggaattaggaccacagagtctac 1139
Db   841 gctggagatcgattttggggaaagaaaataacaggaattaggaccacagagatctac 900

Qy   1140 acagtcgaacttcaactttatgttgatgagtttgtgatgacttccaacaataattc 1199
Db   901 acagtcgaacttcaactttatgttgatgagtttgtgatgacttccaacaataattc 960

Qy   1200 taagtggaagacctataaaggaattgttgaataatgaagaaaagtgttttcagggtaactc 1259
Db   961 taagtggaagacctataaaggaattgttgaataatgaagaaaagtgttttcagggtaactc 1020

Qy   1260 taacttctcgggaccccagtcgcaaacacaattcatcccctccatcgtggccagatatgtcg 1319
Db   1021 taacttctcgggaccccagtcgcaaacacaattcatcccctccatcgtggccagatatgtcg 1080

Qy   1320 ggttgtccccaccagacatggccaccgagggatagcgtcaaggtgggacctccatttggtgca 1379
Db   1081 ggttgtccccaccagacatggccaccgagggatagcgtcaaggtgggacctccatttggtgca 1140

Qy   1380 gattacaacaaggtaatgatcattggtgtggcgtcaagacaagacaagtcaaagcaccagtgttc 1439
Db   1141 gattacaacaaggtaatgatcattggtgtggcgtcaagacaagacaagtcaaagcaccagtgttc 1200

Qy   1440 aactaagaaagaaaggagacaatacaagacgccatctcccctccgaagaaaacatcccacagg 1499
Db   1201 aactaagaaagaaaggagacaatacaagacgccatctcccctccgaagaaaacatcccacagg 1260

Qy   1500 aataaacattacacgtgtcattccttgttgtctcctgtgttgtttcgctgg 1559
Db   1261 aataaacattacacgtgtcattccttgttgtctcctgtgttgtttcgctgg 1320

Qy   1560 aatgggatcttcgcgccttagagaagaagaagaagaagaaagtccgtatggatcagc 1619
Db   1321 aatgggatcttcgcgccttagagaagaagaagaagaagaaagtccgtatggatcagc 1380

Qy   1620 ggaggtccagaaaacgactgttggaaggacagatcaagtaaatatccctttgccagaatcagtc 1679
Db   1381 ggaggtccagaaaacgactgttggaaggacagatcaagtaaatatccctttgccagaatcagtc 1440

Qy   1680 agctgagttacctcagatcgatcatgataatgaagaagaaggaagttgagatccat 1739
Db   1441 agctgagttacctcagatcgatcatgataatgaagaagaaggaagttgagatccat 1500

Qy   1740 cacaagtgtatatggcaggtaa 1761
Db   1501 cacaagtgtatatggcaggtaa 1522

XX
RESULT 5
AC AAA96736
ID AAA96736 standard; DNA; 1871 BP.
XX
AC AAA96736;
XX
DT 19-FEB-2001 (first entry)
XX
DE Polynucleotide isolated from lymph node stromal cells of fsn -/- mice.
XX
KW Lymph node stromal cell; fsn -/- mice; inflammatory disorder;
KW immune system disorder; cancer; viral disorder; HIV infection;
KW blood vessel growth; tumour necrosis factor disorder; arthritis;
KW inflammatory bowel disease; fibroblast growth factor-mediated disorder;
KW cardiac failure; ss.
XX
OS Mus sp.
XX
FH Key   Location/Qualifiers
FT CDS   64..1575
FT   /*tag= a
XX
PN WO200058463-A1.
XX
PD 05-OCT-2000.
XX
PF 18-FEB-2000; 2000WO-W200015.
XX

PR   25-MAR-1999;
PR   26-AUG-1999;
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PI   Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
PI   Murison JG;
XX
WPI; 2000-664924/64.
P-PSDB; AAB19126.
XX

Claim 4: Page 60-61; 75pp; English.

SQ   Sequence 1871 BP; 485 A; 471 C; 484 G; 431 T; 0 other;.

The present sequence represents a polynucleotide sequence which is
isolated from lymph node stromal cells of fsn -/- mice. The
polynucleotides and their polypeptides are useful for treating an
inflammatory disorder, disorder of immune system and cancer selected
from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
viral disorder, in particular HIV infection and for modulating the
growth of blood vessels. The polypeptides are useful for treating a
tumour necrosis factor (TNF) mediated disorder, such as those selected
from arthritis, inflammatory bowel disease and cardiac failure and a
fibroblast growth factor-mediated disorder. It is also useful in assays
to determine biological activity, to raise antibodies, to isolate
cognate corresponding ligand or receptors, as antiinflammatory agents,
corresponding ligands or receptors, to quantify levels of protein or
and in compositions for the treatment of skin, connective tissue and
immune system diseases. The polynucleotide is useful as marker for
tissue, as a chromosome marker or tags in the identification of a
genetic disorder.

CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC

XX
PS   Polypeptide expressed in mammalian fsn -/- lymph node stromal cells,
PT   useful for modulating growth of blood cells, for treating inflammatory
PT   and tumour necrosis factor-mediated disorders, cancer and viral
PT   disorders
XX
DR   P-PSDB; AAB19126.
XX

Query Match   30.6%; Score 539.2; DB 21; Length 1871;
Best Local Similarity   82.8%;
Matches 616; Conservative   0; Mismatches 128; Indels   0; Gaps   0;

Qy   159 cggcggcgcgcgccactggccgcgggctccggcgcggggcctccggcttgctcgccggt 218
            ||||| |||||||||||||||||||||||||||||||||
Db   54 cggcgggaccatcgggggcctgggccggagtcgcctgctgcctgctgctcgccgt 113

Qy   219 ctccgcgccatccggcctccggctgccaggcgaaggagcgggtgggctgatcactagtgac 278
            |||||||||| || ||||||||||||||| ||| ||||||| |||||| |||
Db   114 gtgtgctccgcggcttgcaggcggaagagagcggggtgatcgagctgtggccacatgatgac 173

Qy   279 ttatcaggataggtcacacatgacatctaagaataatatcccggaccccccaatcacac 338
            ||||||||||||| ||||| |||||||||||||| |||||| |||| |||||
Db   174 ctctcaggacagtggcacacatgacatctaaggaagattatccagggactcaccccaattacac 233

Qy   339 tgtttcgcgaaaagacaattacagtacccaaagggaaaagactgattctgaggtgggaga 398
            ||||||| ||||||| ||||||||| |||||| |||| |||||||||| |||||
Db   234 tgtgtgtgaaaagatacatcacagtcccaaaagggaaagaacttattctgaggtgggaga 293

Qy   399 tttggatatcgaatccacagacctgtgctctgactactctctctctctcacacggtctcaga 458
            |||| ||||||||||||||| |||||||||||||| ||||||||||||
Db   294 ttgaacattgagtccaagaaactccgcttctgactatctctcttcagcgagtgcaagaga 353

Qy   459 tcaatatggtccatactgtggaagatgatactgttgtcgtggtgaacaccaag 518
            |||||||||||| |||||||||||||| |||||| | ||||||
Db   354 tcagtatggtccattatgtgggagtggggctgttcccaaagaaacctccggctgaactcaaa 413

Qy   519 tgaagtaacctccgtctttgagggaatgggatcaccaacatttctggccgccggggttttttgtgac 578
            ||||| ||| || ||||||||| |||||| || || |||||||| | |||||
Db   414 cgaagtgactgtctcctctccaagadcggaatctccaattctggccgggggcttctgctgac 473

Qy   579 ctatcgcgagcagccgcaccacagatttaataacctgttcggaacgagctagccaatattt 638
            ||||||||||||||||| |||||||||||||||||||| |||||||||||||||||
Db   474 ctacgccagcagtgaccatccagattaataacctgttcggaacgagctagccaatattt 533

Db   534 cgaggaaaaatcaagacaaatctctgcccagctgttctgaaggacgctacgaggagagcaactc 593

Qy   639 gaagacagaatacagcagaaatctgcccagctgttctgtagagacgttctgaga 698
            || ||| | |||| ||||||||||||||||||||||||||||||| | ||
Db   594 tgggaatacaaagaaggttacagaagacacctcttattgtgcaaaagctgccatccatc 653

Qy   699 tgggaatatgttgagatgatgatctggaagatacctctttattgtgcaaagctgcca 758
            |||||| ||| | |||||||| |||||||||||| ||||||||||||||||||
Db   654 aagggatacacacagaacacctcttattgtgcaaaagctgccatcc 713

Qy   759 aggaataagttgctgatgaactaggtcggtccagatcagtcgtgcttaagcgcaaaggatcag 818
            ||||| |||||||||||||||||||| || ||||| ||||||| |||||| |
Db   714 tcactagaagaaagctctccaagaagcaagccatggctgctacctacccagatcaac 773

Qy   819 tcgatatgaaggagttctgccaatggtgtcttcttcgagggatggttgtctcctgtcaagcaa 878
            |||||||||||||||||||||||||||||| || ||| ||||||||||
Db   774 gcgattctctttttacaaccagg 797

Qy   879 gcgattctctgttacctccaatgg 902
            ||||||||||||||| ||||||||
Db   819 tgcatatgaagggattctggccaatggtgtctccttcgagggatggttctctctgtcaagcaa 878

Claim 4: Page 60-61; 75pp; English.

RESULT   6
AAH34855

ID   AAH34855 standard; cDNA; 2428 BP.
XX
AC   AAH34855;
XX
DT   03-SEP-2001 (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:1937.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999; 99US-0157137.
PR   03-NOV-1999; 99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
WPI; 2001-235357/24.
P-PSDB; AAG75450.
XX
DR   P-PSDB; AAG75450.
XX

Claim 1: Page 3444-3445; 9803pp; English.

PS   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

```
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 2428 BP; 677 A; 560 C; 535 G; 650 T; 6 other;

Query Match             12.6%;  Score 221.8;  DB 22;  Length 2428;
Best Local Similarity   51.8%;
Matches  604;  Conservative   51.8%;  Pred. No. 8.9e-47;
                                                  DB 22;  Mismatches 519;  Indels  42;  Gaps  3;
```

```
Qy    248 agctggggtgtgctctggacactgagacactca...ggatcagatggtggccacactcagcaatcacctca   307
          ||||||||||||||||||||||||||||||| ||||
Db    136 agcaagtcgatggatggtggacaactgctaggagccctagagctggaaccttactcca          195

Qy    308 agaatcatcccggaacatcacccaatcacacatgttttcgaaaagacaattacagtaccaa          367
Db    196 taaactacccacagaacatcccaacagacagctttctgaatgggagatccgtaaaga        255

Qy    368 agggtgaaagaactgattctgaggttggagggagattggatctcgaa---------------  411
Db    256 tgggagagaagagtccgcacaaaattggtgacctgacaattgaaactctgattctgtc     315

Qy    412 --tcccagaccctgttctctgactatctctcttcaccagagtcgacaaatcaatatggtc   469
Db    316 acttcaatttaattactctgaagaatttatatgagtcagcagaacttgaaataggcaaat     375

Qy    470 catactctggaagtactgagttcaaagaaacttcttgttgaacacacaagtgaagtaaccg   529
Db    376 actgtggctctgggggttgcaaatgaaatgcaatcaattcaaaaagacaatgaaatacat   435

Qy    530 tccgctcttgagaggagtgggatcccacacattcctgcgcggggttttttgctgacctatgcgagca   589
Db    436 tgtcgttcactgtggaatccatgttctctgggaccgcggattttttggctccatactctgtta   495

Qy    590 gcggacatccagatttaataacacgttcttggaacgagctagcacnatttattgaagacagaat   649
Db    496 tagataaacaagatctaattactctgttggacactgacactgacactttttggaaaccggagt     555

Qy    650 acagcaaactctgcccagctggttctgtagagacgtagcagcagacattcctgggggatatg   709
Db    556 tcaagtaagtactgccacagctcgtttgtctgctccytttcgctyagatatctggaacattc   615

Qy    710 tagaatggggatatatagagatcacctcttattgtgcaaagctgccatccacatgcagagaatcaattg   769
Db    616 ctcatggaattatatagatcctcgccatgtcatactggctgtgtgcatcagagaagtagtgt   675

Qy    770 ctgatgaactagttggccagtcaagttgcctccagcgccaaagggatccagtcagtatatgag   829
Db    676 caaacacgtggggcccaacaacgctggtttgtctgctccytttgcctyagatatctggaacatc   735

Qy    830 ggattctggccaatggtgtcttcgagggatggtccctgtcaacaagcgattcctgt   889
Db    736 aaataaacagcatcatactgtgtcgagtgactgacccacacagggcaagagaacagttga   795

Qy    890 ttacccccaatcggttcgcagcagatcctgagttttgaacctgacggg-----------c   937
Db    796 ttaagacagctggatctgttatggaacactggggggatctggtgtgtgtgatcgcggatcctc     855

Qy    938 aaatcagagctcttcctccatcagggatggtctccctgtcaacaagttcaactggt   997
Db    856 aaataaacagcatcatactgtgctgagtgactgacccacacagggcaagagaacagttga     915

Qy    998 ctcctggccaagcccaactcaggacgacaagagccacatggtcgcttcggggcgacagtagcn          1057
Db    916 aacccaaaaagccaggctgaaaaaacctggcgctgctttgcactgatg          975

Qy   1058 acaaccacacacgagagtggctggtctgatgtcgatctgggggaagaaagaaaataaacg          1117
Db    976 aatacc--------------agtggttacaaatagattgaataagaagaaataacg          1023
```

```
Qy   1118 gaattaggaccacaggatctacaagtcgaacttcaacttttatgttaagagttttgtga   1177
Db   1024 gcattataaccacctggaatcaccatggtcggagcacaattactgtctgctacagaa    1083

Qy   1178 tgaacttcaaaaaacaataattctaagtcggagaaggaacttatgtcgcttggaataatgaag   1237
Db   1084 tcctgtacagtgatgatgggcagaaaaatggaactgtgtacagatggaaactgccatga   1143

Qy   1238 aaaagtgtttcaaggtaactccaacttttgggaccacggatcgacaaaacaaattcatccctc   1297
Db   1144 ataagatatttcaacaaagatattacagttgtcggtaatacmttttttgcaa        1203

Qy   1298 ccatcctggccagatatgtcggggtttgtccccaacatggcaacaggggatagctggaa   1357
Db   1204 caattattgcacgttttattagagtgaatcctaccaatggcatggcagaaaattgccatga   1263

Qy   1358 agsttggaggtcaatggttgccagatta   1384
Db   1264 aaatggagagctgctggatgtcagtta   1290
```

```
XX
SQ   Sequence 407 BP; 117 A; 83 C; 95 G; 111 T; 1 other;

Query Match             12.5%;  Score 219.4; DB 20; Length 407;
Best Local Similarity   99.5%;  Pred. No. 1.6e-46;
Matches  220;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy  901 ggttgcagcagaatcctgagttttgaacctgacgggcaaatcaggacgcttcctccatgg 960
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183 ggttgcagcagaatcctgagttttgaacctgacgggcaaatcaggacgcttcctccatgg 242

Qy  961 caatgcggtcaatgagatggaggacaccaagttcactgactggccacagagcctcctacg 1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243 caatgcggtcaatgagatggaggacaccaagttcactgactggccacagagcctcctacg 302

Qy 1021 gaccaaggcccatccaatggcttcggggcgacagtagcaacaaccacacaaccacagagtg 1080
        |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  303 gaccaaggcccatccaatggcttcggggcgacagtagcaacaaccacacaaccacagagtgg 362

Qy 1081 ctgggagatcgattcggggagaaaagaaaataaacaggact 1121
        |||||||||||||||||||||||||||||||||||||||||
Db  363 ctgggagatcgattcggggagaaaagaaaataaacaggact 403


RESULT  8

ID   AA251872 standard; cDNA; 2046 BP.
XX
AC   AA251872;
XX
DT   04-JUL-2000  (first entry)
XX
DE   Human Factor 8 Homologue cDNA.
XX
KW   Human; Factor 8 Homologue; F8H; coagulation cofactor; haemostatic;
KW   cerebroprotective; therapeutic; coagulation related disorder;
KW   haemophilia; stroke; screening; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   CDS             19..2028
FT                   /tag= a
FT                   /product= "Factor 8 Homologue"
XX
```

RESULT 9

ID   AAH36736 standard; cDNA; 252 BP.
XX
AC   AAH36736;
XX
DT   03-SEP-2001 (first entry)
XX
DE   Human colon cancer antigen encoding cDNA; 252 BP.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999; 99US-0157137.
PR   03-NOV-1999; 99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
PT   WPI; 2001-213557/24.
PT   P-PSDB; AAG77329.
XX
PT   Claim 1; Page 5674-5675; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated P
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention,
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAH77789 represent sequences used in the exemplification of the
CC   present invention. For example, N.B. Pages 665 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 252 BP; 68 A; 56 C; 68 G; 47 T; 13 other;

Query Match   81.1%; Score 127.2; DB 22; Length 252;
Best Local Similarity 81.1%; Pred. No. 6.9e-23;
Matches 189; Conservative 0; Mismatches 39; Indels 5; Gaps 4;

Qy   1206 gaagacctataaaggaattgtgaataa-tgaagaaaaggtgttcaggtaacctacact 1264
          ||||||||||| |||||| ||| ||||| ||||||| |||||||||||||| |||||||
Db   19 gaggacctatanaggaattgtggtaagtaagaagaataaaaaggtgttcacgggtgaacctacact 78

Qy   1265 tcgggaccacagtgcaaaacaatttcatcccctccatcctggccagatgtgt-gcgggtt 1323
          ||||||||||||| | ||||||||||| |||||||| |||||||||||||||| |||||||
Db   79 tcgggaccacacgtgcacngcagnncactacctcccatngtggccagatatgtgtcgcggatt 138

RESULT 10

ID   AAZ32182 standard; cDNA; 6893 BP.
XX
AC   AAZ32182;
XX
DT   13-JAN-2000 (first entry)
XX
DE   Human coagulation factor V nucleotide sequence.
XX
KW   Human; coding sequence polymorphism; vascular pathology gene;
KW   polymorphic site; phenotype correlation; forensic; paternity testing;
KW   medicine; genetic analysis; vascular disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO9950454-A2.
XX
PD   07-OCT-1999.
XX
PF   26-MAR-1999; 99WO-US06473.
XX
PR   01-APR-1998; 98US-0054272.
XX
PA   (WHED-) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Lander ES, Daley GQ, Cargill M, Ireland JS, Rozen SG;
XX
PT   WPI; 1999-620066/53.
PT   P-PSDB; AAY49563.
XX
PT   Claim 1; Fig 26; 134pp; English.
XX
CC   AAZ32159 to AAZ32194 represent reference alleles for specifically
CC   claimed nucleic acid sequences from the present invention which comprise
CC   polymorphic sites as given in a table in which the nucleotide at the
CC   from 92 single nucleotide polymorphisms in which the nucleotide at the
CC   polymorphic site is different from a nucleotide at the same site in a
CC   reference allele. The nucleic acids, and primers and probes, are used to
CC   identify polymorphisms, which may predispose an individual to disease,
CC   especially a vascular disease. They can also be used in phenotype
CC   correlations, forensics, paternity testing, medicine or genetic
CC   analysis. AAY49550 to AAY49573 represent the proteins which correspond
CC   to some of the reference alleles.
XX
SQ   Sequence 6893 BP; 2090 A; 1700 C; 1423 G; 1680 T; 0 other;

Query Match   51.2%; Score 94.4; DB 20; Length 6893;
Best Local Similarity 51.2%; Pred. No. 8.7e-14;
Matches 249; Conservative 0; Mismatches 231; Indels 6; Gaps 1;

Qy   898 aatggttgcagcagtcctgagtttgaaccgcgacgggcaaatcagagcttcttcctca 957
          ||| ||| | |||||| ||| || ||| || || |||| ||||||||||| ||||
Db   6266 aatggatgttcacacccctcgggtatggaaatggaaaagatagaaacaagacaaatacaca 6325

Qy   958 tgcaactgcgtcaatgagatggtggaccagtccactggtctcctggccaagccccgacct 1017
          ||| ||| |||||||| ||||| ||||||||| |||||||||||| ||||||| |||
Db   6326 gctcttcgttaagaaactcatgtggggaggaattacttgggaaccctccgttccgtccgtctg 6385

Qy   1018 caggacaaggccaccatatggggctcgggcgcacagtagcacaacaacacacacagagag 1077
          ||||||||| ||||| ||||||||||||||| || |||||||||||||||||||||||||
Db   6326 caggacaaggccaccatatggggctcgggcgcacagtagcacaacaacacacagagag 1077

(handwritten) SEQ ID NO: 3
Database: N-Gene Seq - 031802
Ac. No: AAZ 32182

RESULT 11

```
AAT03920
ID   AAT03920 standard; cDNA; 6909 BP.
XX
AC   AAT03920;
XX
DT   20-DEC-1995 (first entry)
XX
DE   Human Factor-V wild-type cDNA.
XX
KW   Factor-V; thrombosis; thrombophilia; diagnosis; anticoagulant;
KW   activated protein-C; APC; homozygosity; heterozygosity; ss.
OS   Homo sapiens.
XX
FH
DE   Bertina RM, Reitsma PH;
XX
PI   WPI; 1995-293134/38.
DR   WO9521938-A1.
XX
PN   17-AUG-1995.
XX
PD   14-FEB-1995;     95WO-EP00553.
PF   14-FEB-1994;     94EP-0200377.
PR   (UYLE-) RIJKSUNIV LEIDEN.
PA
PS   Claim 16; Page 60-65; 98pp; English.
XX
PT   The probes/primers given in AAT03921-31 are used to isolate and/or
PT   amplify and/or detect mutations in human Factor-V DNA (AAT03920),
PT   esp. in DNA encoding the APC binding and/or cleavage site.
CC   Screening for genetic defect associated with thrombosis and/or poor
CC   anticoagulant response to activated protein C - useful to determine
CC   homozygosity or heterozygosity for a mutation in Factor V, Va, VIII
CC   or VIIIa.
SQ   Sequence 6909 BP; 2096 A; 1700 C; 1430 G; 1683 T; 0 other;
```

Query Match 5.3%; Score 92.8; DB 16; Length 6909;
Best Local Similarity 51.0%; Pred. No. 2.3e-13;

```
Qy   6386 aatgcccaagggacgtgtgaatgcctggcaaggcaacacaacaaataagc------ag 6439
            |||||| |||||  |||   ||||  ||| ||         ||            ||
Db   1078 tggctggaggatcgatcttggggggagaaagaaataaacaggaccaaggatct 1137

Qy   6440 tggctggaaattgatctactcaagatcaaggaagaaaccgccaattatacaacaggg 6499
            |||||| |||||||||||||||  |||||||||||  ||||||| |||| |||||
Db   1138 acacagtcgaacttcaacttttattgtttgtgaacttcaaaacaataat 1197

Qy   6500 aagctctctgtcctcgaaatgtatgtaaagaagctaaccatccactacagtgagcaggga 6559
            ||||||||| |||||||||  |||||||| |||||||||||||||| ||||| |||||
Db   1198 tctaagtggagaaacctataaagaattgtgaataatgaagaaaagtgtttcaggtaac 1257

Qy   6560 gtggaatgggaaaccatacagggctgaaatcctccatgtggacagagatttcaaggaaat 6619
            ||| |||||||| |||||| ||||||||||  |||||||||||| ||||||||||||
Db   1258 tctaacttcgggaccccagtgcaaaacaattcatccctccccactcgtggccagatctg 1317

Qy   6620 actaataccaaaggagactgtgaagaacttttccaccccccaatccattccaggttatc 6679
            |||  ||||||| ||||||| |||||||| |||||||| |||| |||||  ||||
Db   1318 cgggttgtgtccccagacaatggcaccaagggggatgagctcatggttgc 1377

Qy   6680 cgtgtcacttcctaaacatggaatcaaagtattgcacttcgcctggaacttcttggctgt 6739
            ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1378 cagatt 1383

Qy   6740 gatatt 6745
            ||||||
Db   6739
```

RESULT 12

```
AAA60448
Matches 248; Conservative 0; Mismatches 232; Indels 6; Gaps 1;

ID   AAA60448 standard; cDNA; 6909 BP.
XX
AC   AAA60448;
XX
DT   09-OCT-2000 (first entry)
XX
DE   Human factor V cDNA sequence SEQ ID NO.26.
XX
KW   Factor V; FV; activated protein C; APC; anticoagulant;
KW   activated protein C resistant factor V; thrombosis; screening;
KW   thrombophilia; ds.
OS   Homo sapiens.
XX
PN   US6066778-A.
XX
PD   23-MAY-2000.
PF   06-NOV-1996;     96US-0746111.
PR   06-NOV-1996;     96US-0746111.
PA   (UNMI ) UNIV MICHIGAN.
PI   Ginsburg D,  Cui J;
DR   WPI; 2000-410682/35.
XX
PT   New transgenic mice expressing activated protein C resistant factor V
PT   and factor V null transgenic mice useful for screening anticoagulants,
PT   as models for human thrombophilia and as models for testing in utero
PT   gene therapy protocols -
```

```
Qy   898 aatcgttgcacagcagatcccttgagtttgaacctgacgcggcaaatcagagagcttctcctca 957
Db  6280 aatggatgttccacaccggtatggaaactgaaga...

Qy  958 tcgacagtccgtccaatgagagtggaaaccaagttcactgtctcctggccagcccgagact 1017
Db  6340 gcttcttcgtttaagaaatcttggtggtgtggaggtcg...

Qy  1018 caggaccaaggcccatatgggcttcgggcgcacgtagta... 1077
Db  6400 aatgcccaagggacgtgtgaatgcctggcaaggcaacacaacaaataagg------ag 6453

Qy  1078 tggctgggatcgattggggggagaaagaaataaacaggaccaaggatct 1137
Db  6454 actaataccaaaggacactgaagaacttttccacccccaatccattccaggttatc 6513

Qy  1138 acacagtcgaacttttattgtttgtgaacttcactgtctcaaaacaataat 1197
Db  6514 aagctctctgtcctcgaaatgtataaagaagctaaccatccactacagtgagcaggga 6573

Qy  1198 tctaagtggagaacctataaagaattgtgaataatgaagaaaagtgtttcaggtaac 1257
Db  6574 gtggaatgggaaaccatacagggctgaaatcctccatgtggacagagatttcaaggaaat 6633

Qy  1258 tctaacttcgggaccccagtcaaaacaattcatccctccccactcgtggccagatctgtg 1317
Db  6634 actaataccaaaggagactgtgaagaacttttccaccccccaatccattccagg 6693

Qy  1318 cgggttgtgtccccagacaatggcaccaagggggatgagctcatggttgc 1377
Db  6694 cgtgtcacttcctaaacatggaatcaaagtattacacttcgcctggaacttcttggctgt 6753

Qy  1378 cagatt 1383
Db  6754 gatatt 6759
```

XX
PS Example 1; Column 77-84; 76pp; English.
XX
CC The present invention describes transgenic mice (I) and (II) containing
CC modifications in the factor V gene, where (I) expresses an activated
CC protein C (APC) resistant factor V and (II) lacks the ability to express
CC wild-type factor V. The transgenic animals (I) and (II) are useful for
CC screening compounds with anticoagulant activity. Methods from the present
CC invention, and the transgenic animals, are also useful in providing
CC models for human thrombophilia. These models are useful in providing
CC insight into the basic regulatory mechanisms of blood coagulation and
CC pathogenesis of human thrombosis. In addition, factor V null transgenic
CC mice, especially pregnant females may be used as a model system to test
CC in utero gene replacement therapy protocols. The present sequence
CC represents a human factor V cDNA sequence, which is used in an example
CC from the present invention.
XX
SQ Sequence 6909 BP; 2096 A; 1701 C; 1430 G; 1682 T; 0 other;

Query Match   5.3%; Score 92.8; DB 21; Length 6909;
Best Local Similarity   51.0%; Pred. No. 2.3e-13;
Matches 248; Conservative   0; Mismatches 232; Indels   6; Gaps   1;

Qy   898 aatggttgcagcagatcctgagtcttgaacctgacgggcaaatcagagctttctcctca 957
            |||| || |||||| || | ||||||| |||| || || |||||| ||| ||||||
Db   6280 aatggagtgttccacacccctgggtatggaaatgaaacaagacaaatcaca 6339

Qy   958 tggcagtcgtcaatgaagtgagagaccaagttcactgttctcctggccaagccgcagcaacttcactgttctcctggccaagccgcagcaactt 1017
            |||| | | | | || | || | || | ||| || | ||| ||| |||| |||
Db   6340 gcttcttcgtttaagaaatcttggtcggtgggaggaattactgggaagcacctttcgtgccgctctg 6399

Qy   1018 caggaccaagccgccatcgggttcgggccacagtagcacaaccacacacacagagag 1077
            |||| | || | | | | || |||| | | || || || ||| | |||
Db   6400 aatgcccaggacgtgtgaatgcctggcaagccaagcaacaaatagc------ag 6453

Qy   1078 tggctggaatcgattcgggggagaaaaagaaaaatacacggaataagaaccaccaggatct 1137
            ||| ||| || || || ||| ||| ||||||||| ||| || |||||| ||| |||
Db   6454 tggctagaaattgatactccaagatcaagaagataacgcaatataacactaacaggctgc 6513

Qy   1138 acacagtcgaactcaacttttatgttaagaagttgtcatgaactccaaaacacatat 1197
            ||| | || |||||| | | || | | || | || || | | || | | |
Db   6514 aagtctgtctctgaaatgtatgttaagagctatacacctaacactggtcatcacagga 6573

Qy   1198 tctaagtggaagaccctatataaaggaagttgtgaataatgaagaaaagagt 1257
            | || || || ||| | | | |||| ||| | | | || ||||||||||
Db   6574 gtgggaagtggaacatacaggcgcaaagccaagcaagccaagccacaggtaac 6633

Qy   1258 tctacttccggaccagcaactcctccccatcctggccagatatgtg 1317
            ||| | |||||||| | | || ||| |||| |||| ||| | || || |
Db   6634 actaataccaaaggacatgtgaagaacttccaaccccaatcattccaaggttatc 6693

Qy   1318 cgggttgtccccagaacatggcaccagagatagcctgaagtgagctcaatggttgc 1377
            ||||| | | || ||| ||| | ||| | | || || || ||||||||||| |
Db   6694 cgtgtcattcctaaaacacatgtgaatacaagtattacacttcgcctggaacttttgctgt 6753

Qy   1378 cagatt 1383
            ||||||
Db   6754 gatatt 6759

RESULT 13
XX
AA AZ39631
ID AZ39631 standard; cDNA, 6909 BP.
XX
AC AZ39631;
XX
DT 28-FEB-2000 (first entry)
XX
DE Human Vth aggregation factor protein encoding cDNA.
XX
KW Gene polymorphism; human; Vth aggregation factor protein encoding cDNA;
KW diabetes; PPCR; SSCP; fluorescence-based polymerase chain reaction;

KW single strand conformation polymorphism; ss.
XX
OS Homo sapiens.
XX
CC JP11313676-A.
XX
PN 16-NOV-1999.
XX
PN 30-APR-1998;   98JP-0120217.
XX
PR 30-APR-1998;   98JP-0120217.
XX
PA (SAKA) OTSUKA PHARM CO LTD.
XX
PT Discrimination of human V aggregation factor gene polymorphism -
XX
PT WPI; 2000-057352/05.
XX
PS Discrimination; Page 13-15; 34pp; Japanese.
XX
CC The invention provides a method for the discrimination of the gene
CC polymorphism of human Vth aggregation factor, where one of the following
CC (1) to (6) residues/nucleotides in the aggregation gene is discriminated
CC in the patient to be tested :(1) residue 495: guanine (G) or adenine (A),
CC (2) residue 642: (G) or thymine (T), (3) residue 2663: (G) or (A), (4)
CC residue 2763: (G) or (A), (5) residue 2863: (A) or (G), (6) residue 5112:
CC (A) or (G). The method is useful in the genetic diagnosis of a diabetes
CC patient. The method uses FPCR-SSCP (fluorescence-based polymerase chain
CC reaction-single strand conformation polymorphism) for analyzing DNA
CC samples for polymorphisms. The present sequence represents a cDNA
CC encoding a human Vth aggregation factor protein.
XX
SQ Sequence 6909 BP; 2096 A; 1701 C; 1430 G; 1682 T; 0 other;

Query Match   5.3%; Score 92.8; DB 21; Length 6909;
Best Local Similarity   51.0%; Pred. No. 2.3e-13;
Matches 248; Conservative   0; Mismatches 232; Indels   6; Gaps   1;

Qy   898 aatggttgcagcagatcctgagtcttgaacctgacgggcaaatcagagctttctcctca 957
            |||| || |||||| || | ||||||| |||| || || |||||| ||| ||||||
Db   6280 aatggagtgttccacacccctgggtatggaaatgaaacaagacaaatcaca 6339

Qy   958 tggcagtcgtcaatgaagtgagagaccaagttcactgttctcctggccaagccgcagcaactt 1017
            |||| | | | | || | || | || | ||| || | ||| ||| |||| |||
Db   6340 gcttcttcgtttaagaaatcttggtcggtgggaggaattactgggaagcacctttcgtgccgctctg 6399

Qy   1018 caggaccaagccgccatcgggttcgggccacagtagcacaaccacacacacagagag 1077
            |||| | || | | | | || |||| | | || || || ||| | |||
Db   6400 aatgcccaggacgtgtgaatgcctggcaagccaagcaacaaatagc------ag 6453

Qy   1078 tggctggaatcgattcgggggagaaaaagaaaaatacacggaataagaaccaccaggatct 1137
            ||| ||| || || || ||| ||| ||||||||| ||| || |||||| ||| |||
Db   6454 tggctagaaattgatactccaagatcaagaagataacgcaatataacactaacaggctgc 6513

Qy   1138 acacagtcgaactcaacttttatgttaagaagttgtcatgaactccaaaacacatat 1197
            ||| | || |||||| | | || | | || | || || | | || | | |
Db   6514 aagtctgtctctgaaatgtatgttaagagctatacacctaacactggtcatcacagga 6573

Qy   1198 tctaagtggaagaccctatataaaggaagttgtgaataatgaagaaaagagt 1257
            | || || || ||| | | | |||| ||| | | | || ||||||||||
Db   6574 gtgggaagtggaacatacaggcgcaaagccaagcaagccaagccacaggtaac 6633

Qy   1258 tctacttccggaccagcaactcctccccatcctggccagatatgtg 1317
            ||| | |||||||| | | || ||| |||| |||| ||| | || || |
Db   6634 actaataccaaaggacatgtgaagaacttccaaccccaatcattccaaggttatc 6693

Qy   1318 cgggttgtccccagaacatggcaccagagatagcctgaagtgagctcaatggttgc 1377
            ||||| | | || ||| ||| | ||| | | || || || ||||||||||| |
Db   6694 cgtgtcattcctaaaacacatgtgaatacaagtattacacttcgcctggaacttttgctgt 6753

Qy   1378 cagatt 1383
            ||||||
Db   6754 gatatt 6759

Db   6754 gatatt 6759

RESULT 14
ID   AAD07368
     AAD07368 standard; DNA; 6909 BP.
XX
AC   AAD07368;
XX
DT   10-AUG-2001 (first entry)
XX
DE   Human DNA encoding factor V protein (marker 19).
XX
KW   Human; cytostatic; diagnosis; prostate disorder; prostate cancer; BPH;
KW   benign prostatic hyperplasia; factor V protein; foetal liver protein;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200136674-A2.
XX
PD   25-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-GB04267.
XX
PR   13-NOV-1999; 99GB-0026805.
XX
PA   (ASTR) ASTRAZENECA AB.
PA   (ASTR) ASTRAZENECA UK LTD.
XX
PI   Bull JH, Ellison G, Paskins LD;
XX
DR   WPI; 2001-343837/36.
XX
PT   Claim 1; Page 63-64; 69pp; English.
XX
PT   Diagnosing and monitoring prostate disorders, by analysis of 26 gene
PT   transcripts that exhibit aberrant expression levels in prostate
PT   disorder tissues, and provides a means of early diagnosis -
XX
CC   The patent discloses a method for diagnosing, prognosing or monitoring
CC   a prostate disorder which involves the analysis of 26 gene transcripts
CC   (referred as markers) that exhibit aberrant expression levels in prostate
CC   disorder tissues and provides a means of early diagnosis. This method is
CC   useful for diagnosing, prognosing or monitoring a prostate disorder. It
CC   also provides a means of distinguishing prostate cancer from benign
CC   prostatic hyperplasia (BPH) and for identifying potential anti-prostate
CC   disorder therapeutic compounds. The present sequence is a human DNA encoding factor V protein (referred
CC   as marker 19). The factor V protein is identified as a foetal liver
CC   protein.
XX
SQ   Sequence 6909 BP; 2096 A; 1701 C; 1430 G; 1682 T; 0 other;

Query Match   5.3%;
Best Local Similarity   51.0%;
Matches 248; Conservative   0; Mismatches 232; Indels   6; Gaps   1;

Qy   898 aatggtgcagcagatcctgagtttgaacctgacgggcaaatcagaagctctcctca 957
                ||||| | ||| | | |||| | ||| |||||||||||| |||
Db   6280 aatggagtgttccacaccccgggtatgagaaatggaaagatggaaaacaaggcaaatcaca 6339

Qy   958 tggcacgtcggtcaatgagagaccaagttcactggtctcctggccaagccaccactt 1017
                |||||| || ||| | | || | ||||| ||||||||||||
Db   6340 gctctctcgtttaagaaacttcggtctgggggagttactgggaaccttcctgccctgtctg 6399

Qy   1018 caggacccaagcccatcaggcttcggcgcgacactagcacaacaacaacaacgagag 1077
                | | | |||| | || || ||| | | || |||||||||||| |||
Db   6400 aatgcccaaggcgacgtgtgaatgcctgcaagccaagccaaacaacaataagc-----ag 6453

Qy   1078 tggctggagatcggatctgggggagaaagaaaatcaacaggaaataacaggacaacgatct 1137
                ||||| | | | ||| ||||| | ||| | | || | | || ||
Db   6400 ...

Db   6454 tggctagaaattgatctactcaaggatcaagaagaagtaacggcaaattatacacaaggcgc 6513
               | || || |||| |||| ||| || |||| || || | ||
Qy   1138 acacagtcgaacttcaacttttatgttaagagttttgtgtggatgaacttcaaaacaataat 1197
Db   6514 aagtctctgtcctctgaaatgtatgtaaagagctataccatccacctacagtgagtcagga 6573
               | || | | | |||| ||| | ||| | || || | || || || ||
Qy   1198 tctaagtggaagacctataaagagaattgtcgaataatgaagaaaaagtgttcaaggtaac 1257
Db   6574 gtggaaagtggaaacatacaggctgaaatccctccatgtggacaagactttttgaaggaaat 6633
               || ||||| | | | || || | || || || | | ||| |||| ||||||
Qy   1258 tctaacttctgggcaccagtcgcaaacaacaatttcatcccctccactgtggccagatcgtg 1317
Db   6634 actaataccaaggacgctgaagaaccttttccaacccoccaccaatcattccaggcttatc 6693
               || | || ||| | ||||||| | | | | || || || |||| | || | |||
Qy   1318 cgggttgtccccaggactggcacccgaggagtagagagacctttgaaggtgggaaggccattggttgc 1377
Db   6694 cgtcgtcatccttaaaaacatggaatcaaaagtattacactcgcctgaaacttggctgt 6753
               || | || || ||| ||| | ||||| | || || | | ||| || |||||
Qy   1378 cagatt 1383
Db   6754 gatatt 6759

RESULT 15
ID   AAT33948
     AAT33948 standard; cDNA; 6925 BP.
XX
AC   AAT33948;
XX
DT   14-DEC-1996 (first entry)
XX
DE   Human Factor V mutant cDNA.
XX
KW   Factor V; activated Protein C resistance; APC; genetic screening;
KW   allele; point mutation; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   allele   1691
FT   /tag= a
FT   /note= "base 1691 is guanine in the normal allele"
XX
PN   WO9630546-A1.
XX
PD   03-OCT-1996.
XX
PF   22-MAR-1996; 96WO-US003881.
XX
PR   24-MAR-1995; 95US-0410488.
XX
PA   (SCRI) SCRIPPS RES INST.
XX
PI   Gandrille S, Greengard J, Griffin JH;
XX
DR   WPI; 1996-455389/45.
XX
PT   Claim 22; Page 146-149; 175pp; English.
XX
PT   Detection of Factor V gene mutation - by PCR amplification to
PT   identify exon 10 guanine 205 or 1691 to adenine substitution, which
PT   results in activated Protein C resistance
XX
CC   A mutant human Factor V cDNA (AAT33948) has adenine at position
CC   1691,whereas the normal cDNA (AAT33942) has guanine at this
CC   position. The substn. results in activated Protein C resistance
CC   in humans. The mutation can be detected by PCR amplification
CC   of the affected region (see also AAT33935, AAT33939-41) and
CC   analysis of the products (see also AAT33949-50), either by nucleotide
CC   sequencing or by differential digestion using MnlI restriction
CC   endonuclease (see also AAT33937-38).
XX
SQ   Sequence 6925 BP; 2106 A; 1703 C; 1432 G; 1684 T; 0 other;

```
Query Match      5.3%;  Score 92.8;  DB 17;  Length 6925;
Best Local Similarity   51.0%;  Pred. No. 2.3e-13;
Matches  248;  Conservative   0;  Mismatches  232;  Indels    6;  Gaps    1;

Qy    898 aatggttgcagcagatccttgagtttttgaacctgacggtgcaaatcagagcttcctca  957
          ||||| ||  || || || || ||| | || ||
Db   6280 aatggatgttccacocctgggtatggaaaatggaaagatagaaaacaagcaaatacaca  6339

Qy    958 tggcagtcggtcaatggagtggagaccaagttcactggtctcctggccaagccggactt  1017
          ||  |     | ||| | || ||| |  |||| || |||| || ||| || || | || |
Db   6340 gctcttccgttaagaaatcctcggtgtggggagattactgggaaccottccgtgccgtctg  6399

Qy   1018 caggaccaaggcccatcatggcttcgggcgacgtagtcacaacaccaaaccacgagag  1077
          | |  || ||     |||| ||   ||  || ||  |  ||| ||   ||  |   || |
Db   6400 aatgcccagggacgtgtgaatgcctggcaagccaagcaaacaacataagc-------ag  6453

Qy   1078 tggctggagatcgatttcgggggagaaaaagaaaataacaggaattaggaccacaggatct  1137
          ||  |   ||| | ||| || || || ||  | | || ||  || || || |  || ||
Db   6454 tggctagaaattgatctaactcaagatcaagaagataacggcaaatatataacacagggctgc  6513

Qy   1138 acacagtcgaacttcaacttttatgttaagagtttgatgaacttcaaaaacaataat  1197
          | |     |   | ||| |  |  ||| ||| ||||  | ||| |  || | ||
Db   6514 aagtctgtccctgaaatgtatgtaaagctataacatccactacagtgagcaggga  6573

Qy   1198 tctaagtgggaagacctataaaggaattgtgaataatgaagaaaaggtgttccagggtaac  1257
          |     |  ||| ||| ||   || || || ||| || ||   || |  || || ||
Db   6574 gtggaatggaaaccatacagcgctgaaatccctcctggtggacaagattttgaaggaaat  6633

Qy   1258 tctaacttctcgggaccccagtgtgcaaaacaattcatccctccccatcgtggccagatatgtg  1317
          ||||          |       ||     |  || || ||| ||| || || ||
Db   6634 actaataccaaaaggacatgtgaagaactttttcaacccccaatcattccaggttctttctc  6693

Qy   1318 cggggttgtcccccagacatggcaccagagaggatgtagcttgaaggtggagctcaattggttgc  1377
          |  |   ||| |  || |  |  ||  | || ||   ||| ||||  | | ||| || | ||
Db   6694 cgtgtcattcctaaaacatggaatcaaagtattacacttcgcctggaactccttggctgt  6753

Qy   1378 cagatc  1383
          |  | |
Db   6754 gatatt  6759
```

```
=> fil reg
COST IN U.S. DOLLARS                    SINCE FILE      TOTAL
                                          ENTRY        SESSION
FULL ESTIMATED COST                       0.38          2.90

FILE 'REGISTRY' ENTERED AT 15:46:21 ON 09 JUL 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 American Chemical Society (ACS)

STRUCTURE FILE UPDATES:    8 JUL 2002  HIGHEST RN 437701-77-4
DICTIONARY FILE UPDATES:   8 JUL 2002  HIGHEST RN 437701-77-4

TSCA INFORMATION NOW CURRENT THROUGH January 7, 2002

   Please note that search-term pricing does apply when
   conducting SmartSELECT searches.

Crossover limits have been increased.  See HELP CROSSOVER for details.

Calculated physical property data is now available.  See HELP PROPERTIES
for more information.  See STNote 27, Searching Properties in the CAS
Registry File, for complete details:
http://www.cas.org/ONLINE/STN/STNOTES/stnotes27.pdf

=> e animal cub domain protein/cn 5
E1          1      ANIMAL BLACK/CN
E2          1      ANIMAL BLOOD, DENATURED/CN
E3          0 -->  ANIMAL CUB DOMAIN PROTEIN/CN
E4          1      ANIMAL EXTRACTS, LIVER/CN
E5          1      ANIMAL FAT OLEINS/CN

=> e cub domain/cn 5
E1          1      CUAUII/CN
E2          1      CUAUTHEMONE/CN
E3          0 -->  CUB DOMAIN/CN
E4          1      CUB DOMAIN-CONTAINING PROTEIN 1 (HUMAN A549/CALU 6 CELL GENE
                     CDCP1 PRECURSOR)/CN
E5          1      CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA)/CN

=> e
E6          1      CUB2/CN
E7          1      CUBA BLACK/CN
E8          1      CUBA ORANGE/CN
E9          1      CUBA6/CN
E10         1      CUBANA BROWN/CN
E11         1      CUBANA BROWN R/CN
E12         1      CUBANE/CN
E13         1      CUBANE RADICAL CATION/CN
E14         1      CUBANE-1,2,4,7-TETRAAMMONIUM DINITRAMIDE/CN
E15         1      CUBANE-1,4-DIAMMONIUM DINITRAMIDE/CN
E16         1      CUBANE-1,4-DICARBOXYLIC ACID-1,4-PHENYLENEDIAMINE COPOLYMER/
                     CN
E17         1      CUBANE-1,4-DICARBOXYLIC ACID-1,4-PHENYLENEDIAMINE COPOLYMER,
                     SRU/CN

=> s e4-5
            1 "CUB DOMAIN-CONTAINING PROTEIN 1 (HUMAN A549/CALU 6 CELL GENE
              CDCP1 PRECURSOR)"/CN
            1 "CUB-DOMAIN-CONTAINING PROTEIN RP43 (RIFTIA PACHYPTILA)"/CN
L1          2 ("CUB DOMAIN-CONTAINING PROTEIN 1 (HUMAN A549/CALU 6 CELL GENE
              CDCP1 PRECURSOR)"/CN OR "CUB-DOMAIN-CONTAINING PROTEIN RP43

Searched by: Mary Hale 308-4258 CM-1 1E01
```

```
                    (RIFTIA PACHYPTILA)"/CN)

   => e coagulation factors v/cn 5
   E1          1      COAGULATION FACTOR XIA/CN
   E2          1      COAGULATION FACTORS (BLOOD)/CN
   E3          0 --> COAGULATION FACTORS V/CN
   E4          1      COAGULEN/CN
   E5          3      COAGULIN/CN

   => e coagulation factors xiii/cn 5
   E1          1      COAGULATION FACTOR XIA/CN
   E2          1      COAGULATION FACTORS (BLOOD)/CN
   E3          0 --> COAGULATION FACTORS XIII/CN
   E4          1      COAGULEN/CN
   E5          3      COAGULIN/CN

   => e milk fat globule egf factor 8/cn 5
   E1          1      MILK BUNDLE-1 PROTEIN (SYNTHETIC)/CN
   E2          1      MILK FAT GLOBULE ANTIGEN NP5 (HUMAN FRAGMENT)/CN
   E3          0 --> MILK FAT GLOBULE EGF FACTOR 8/CN
   E4          1      MILK GLASS/CN
   E5          1      MILK GROWTH FACTOR/CN

   => e transcriptional repressor/cn 5
   E1          1      TRANSCRIPTIONAL REGULATORYPROTEIN (YERSINIA PESTIS STRAIN CO
                      92 GENE YPO1938)/CN
   E2          1      TRANSCRIPTIONAL REPRESSION OF HYC AND HYP OPERONS (ESCHERICH
                      IA COLI O157:H7 STRAIN EDL933 GENE HYCA)/CN
   E3          0 --> TRANSCRIPTIONAL REPRESSOR/CN
   E4          1      TRANSCRIPTIONAL REPRESSOR (AEDES AEGYPTI STRAIN UGAL REDUCED
                      )/CN
   E5          1      TRANSCRIPTIONAL REPRESSOR (AGROBACTERIUM TUMEFACIENS STRAIN
                      C58 GENE ACCR)/CN

   => e
   E6          1      TRANSCRIPTIONAL REPRESSOR (AGROBACTERIUM TUMEFACIENS STRAIN
                      C58 GENE TRAM)/CN
   E7          1      TRANSCRIPTIONAL REPRESSOR (ARABIDOPSIS THALIANA CLONE G11C3
                      GENE RPD3A)/CN
   E8          1      TRANSCRIPTIONAL REPRESSOR (ASPERGILLUS ORYZAE STRAIN TK3 GEN
                      E CREA)/CN
   E9          1      TRANSCRIPTIONAL REPRESSOR (BACILLUS HALODURANS STRAIN C-125
                      GENE BH0826)/CN
   E10         1      TRANSCRIPTIONAL REPRESSOR (BACILLUS HALODURANS STRAIN C-125
                      GENE BH0847)/CN
   E11         1      TRANSCRIPTIONAL REPRESSOR (BACTERIOPHAGE .PHI.CH1)/CN
   E12         1      TRANSCRIPTIONAL REPRESSOR (BETA-GALACTOSIDASE GENE) (BACILLU
                      S HALODURANS STRAIN C-125 GENE LACR)/CN
   E13         1      TRANSCRIPTIONAL REPRESSOR (BRUCELLA MELITENSIS STRAIN 16M GE
                      NE BMEI1046)/CN
   E14         1      TRANSCRIPTIONAL REPRESSOR (BRUCELLA MELITENSIS STRAIN 2308 G
                      ENE ERID)/CN
   E15         1      TRANSCRIPTIONAL REPRESSOR (BUTTERFLY GENE EXTRA SEX COMBS)/C
                      N
   E16         1      TRANSCRIPTIONAL REPRESSOR (CANDIDA ALBICANS CLONE P371 GENE
                      TUP1 REDUCED)/CN
   E17         1      TRANSCRIPTIONAL REPRESSOR (CANDIDA UTILIS STRAIN NRRL Y1084
                      GENE MIG1)/CN

   => s transcriptional repressor ?/cn
   L2         171 TRANSCRIPTIONAL REPRESSOR ?/CN


Searched by: Mary Hale 308-4258 CM-1 1E01
```

```
=> e "ae-binding protein-1"/cn 5
E1          1       AE-100/CN
E2          1       AE-3/CN
E3          0 --> AE-BINDING PROTEIN-1/CN
E4          1       AE-CELLULOSE/CN
E5          1       AE-F 117233/CN

=> e neurophilins 1/cn 5
E1          1       NEUROPEPTIDE, SUBOESOPHAGEAL GANGLION (BOMBYX MORI BOMPBAN G
                    AMMA-)/CN
E2          1       NEUROPEPTIDE-GLUTAMIC ACID-ISOLEUCINE/CN
E3          0 --> NEUROPHILINS 1/CN
E4          1       NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-1 REDUCED)/CN
E5          1       NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-4 REDUCED)/CN

=> e neurophilins 2/cn 5
E1          1       NEUROPEPTIDE, SUBOESOPHAGEAL GANGLION (BOMBYX MORI BOMPBAN G
                    AMMA-)/CN
E2          1       NEUROPEPTIDE-GLUTAMIC ACID-ISOLEUCINE/CN
E3          0 --> NEUROPHILINS 2/CN
E4          1       NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-1 REDUCED)/CN
E5          1       NEUROPHYSIN (BUFO JAPONICUS CLONE PVT-4 REDUCED)/CN

=> e discoidin domain/cn 5
E1          1       DISCOGUATTINE/CN
E2          1       DISCOIDIC ACID/CN
E3          0 --> DISCOIDIN DOMAIN/CN
E4          1       DISCOIDIN DOMAIN RECEPTOR 1 KINASE/CN
E5          1       DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYM
                    E DDR1D C-TERMINAL FRAGMENT)/CN

=> e
E6          1       DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYM
                    E DDR1E FRAGMENT)/CN
E7          1       DISCOIDIN DOMAIN RECEPTOR 2 KINASE/CN
E8          1       DISCOIDIN DOMAIN RECEPTOR KINASE (HUMAN GENE CAK ISOENZYME C
                    AK II)/CN
E9          1       DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
                    GENE CAK ISOENZYME CAK I)/CN
E10         1       DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
                    GENE CAK ISOENZYME CAK II)/CN
E11         1       DISCOIDIN DOMAIN RECEPTOR TYROSINE KINASE/CN
E12         1       DISCOIDIN DP1 (DICTYOSTELIUM DISCOIDEUM STRAIN AX2)/CN
E13         1       DISCOIDIN I-A (DICTYOSTELIUM DISCOIDEUM REDUCED)/CN
E14         1       DISCOIDIN I-C (DICTYOSTELIUM DISCOIDEUM REDUCED)/CN
E15         1       DISCOIDIN II (DICTYOSTELIUM DISCOIDEUM CLONE D2)/CN
E16         1       DISCOKIOLIDE A/CN
E17         1       DISCOKIOLIDE B/CN

=> s e4-11
                    1 "DISCOIDIN DOMAIN RECEPTOR 1 KINASE"/CN
                    1 "DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYME
                      DDR1D C-TERMINAL FRAGMENT)"/CN
                    1 "DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYME
                      DDR1E FRAGMENT)"/CN
                    1 "DISCOIDIN DOMAIN RECEPTOR 2 KINASE"/CN
                    1 "DISCOIDIN DOMAIN RECEPTOR KINASE (HUMAN GENE CAK ISOENZYME CAK
                      II)"/CN
                    1 "DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
                      GENE CAK ISOENZYME CAK I)"/CN
                    1 "DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
                      GENE CAK ISOENZYME CAK II)"/CN

Searched by: Mary Hale 308-4258 CM-1 1E01
```

```
                       1  "DISCOIDIN DOMAIN RECEPTOR TYROSINE KINASE"/CN
    L3                 8  ("DISCOIDIN DOMAIN RECEPTOR 1 KINASE"/CN OR "DISCOIDIN DOMAIN
                          RECEPTOR 1 KINASE (HUMAN GENE DDR1 ISOENZYME DDR1D C-TERMINAL
                          FRAGMENT)"/CN OR "DISCOIDIN DOMAIN RECEPTOR 1 KINASE (HUMAN
                          GENE DDR1 ISOENZYME DDR1E FRAGMENT)"/CN OR "DISCOIDIN DOMAIN
                          RECEPTOR 2 KINASE"/CN OR "DISCOIDIN DOMAIN RECEPTOR KINASE (HUMA
                          N GENE CAK ISOENZYME CAK II)"/CN OR "DISCOIDIN DOMAIN RECEPTOR
                          KINASE (MUS MUSCULUS STRAIN C57BL GENE CAK ISOENZYME CAK I)"/CN
                          OR "DISCOIDIN DOMAIN RECEPTOR KINASE (MUS MUSCULUS STRAIN C57BL
                          GENE CAK ISOENZYME CAK II)"/CN OR "DISCOIDIN DOMAIN RECEPTOR
                          TYROSINE KINASE"/CN)

=> e nhp/cn 5
E1            1     NHJ 44/CN
E2            1     NHMO/CN
E3            0  --> NHP/CN
E4            1     NHP2/RS6-LIKE PROTEIN (TRYPANOSOMA BRUCEI STRAIN 427 GENE NH
                   P2)/CN
E5            1     NHPTT/CN

=> fil medl,hcap,biosis,embase,jicst,wpids
COST IN U.S. DOLLARS                              SINCE FILE      TOTAL
                                                    ENTRY        SESSION
FULL ESTIMATED COST                                 46.28         49.18


FILE 'MEDLINE' ENTERED AT 15:49:39 ON 09 JUL 2002

FILE 'HCAPLUS' ENTERED AT 15:49:39 ON 09 JUL 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'BIOSIS' ENTERED AT 15:49:39 ON 09 JUL 2002
COPYRIGHT (C) 2002 BIOLOGICAL ABSTRACTS INC.(R)

FILE 'EMBASE' ENTERED AT 15:49:39 ON 09 JUL 2002
COPYRIGHT (C) 2002 Elsevier Science B.V.  All rights reserved.

FILE 'JICST-EPLUS' ENTERED AT 15:49:39 ON 09 JUL 2002
COPYRIGHT (C) 2002 Japan Science and Technology Corporation (JST)

FILE 'WPIDS' ENTERED AT 15:49:39 ON 09 JUL 2002
COPYRIGHT (C) 2002 THOMSON DERWENT

=> s (l1 or cub domain protein) and (coagulat? factor? or blood coagulation or milk
fat globule egf factor 8 or mfg human milk or transcriptional repressor or crtr-1
protein or ae-binding protein 1 or neurophilin?)
L4            0 FILE MEDLINE
L5            1 FILE HCAPLUS
L6            0 FILE BIOSIS
L7            0 FILE EMBASE
L8            0 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'
L9            1 FILE WPIDS

TOTAL FOR ALL FILES
L10           2 (L1 OR CUB DOMAIN PROTEIN) AND (COAGULAT? FACTOR? OR BLOOD COAGU


Searched by: Mary Hale 308-4258 CM-1 1E01
```

LATION OR MILK FAT GLOBULE EGF FACTOR 8 OR MFG HUMAN MILK OR
TRANSCRIPTIONAL REPRESSOR OR CRTR-1 PROTEIN OR AE-BINDING PROTEI
N 1 OR NEUROPHILIN?)

=> dup rem l10
PROCESSING COMPLETED FOR L10
L11   1 DUP REM L10 (1 DUPLICATE REMOVED)

=> d cbib abs

L11   ANSWER 1 OF 1   HCAPLUS   COPYRIGHT 2002 ACS   DUPLICATE 1
2001:300872   Document No. 134:321614   Protein and cDNA sequences of a novel
      human CUB-domain containing protein and uses thereof in diagnosis, therapy
      and drug screening.   Donoho, Gregory; Turner, C. Alexander, Jr.; Nehls,
      Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon
      Genetics Incorporated, USA).   PCT Int. Appl. WO 2001029219 A1 20010426, 33
      pp.   DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR,
      BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE,
      GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,
      LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD,
      SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM,
      AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM,
      CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT,
      SE, SN, TD, TG.   (English).   CODEN: PIXXD2.   APPLICATION: WO 2000-US28798
      20001008.   PRIORITY: US 1999-PV160285 19991019; US 2000-PV183583 20000218.
AB   This invention provides protein and cDNA sequences for newly identified
      human proteins, designated NHPs, which shares structural similarity with
      animal **CUB domain proteins**,
      coagulation factors V and XIII, **milk**
      fat globule-EGF factor 8,
      5 **transcriptional repressor AE-**
      binding protein-1, and neuropilins 1 and 2
      (which, like the presently described protein, contain both CUB and
      discoidin domains).   While NHP shares sequence homol. with other
      CUB-domain contg. proteins, its primary sequence is unique.   Its
      expression is detected in human cell lines, and human prostate, pituitary,
      fetal brain, brain, thymus, spleen, lymph node, trachea, kidney, fetal
      liver, thyroid, adrenal gland, salivary gland, stomach, small intestine,
      colon, muscle, heart, mammary gland, adipose, skin, esophagus, bladder,
      cervix, rectum, and testis cells.   In one embodiment, the invention
      relates to diagnostic assays for detecting diseases assocd. with
      inappropriate NHP activity or levels.   Also disclosed are methods for
      utilizing NHP in drug screening assays and in therapy directed against
      diseases assocd. with inappropriate NHP activity or levels.

=> s (l1 or cub domain protein) and (discoidin domain or l3 or discoidin receptor
or ddr2 gene product?)
L12   0 FILE MEDLINE
L13   ·1 FILE HCAPLUS
L14   0 FILE BIOSIS
L15   0 FILE EMBASE
L16   0 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

Searched by: Mary Hale 308-4258 CM-1 1E01

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'
L17   0 FILE WPIDS

TOTAL FOR ALL FILES
L18   1 (L1 OR CUB DOMAIN PROTEIN) AND (DISCOIDIN DOMAIN OR L3 OR DISCOI
             DIN RECEPTOR OR DDR2 GENE PRODUCT?)

=> s l18 not l10
L19   0 FILE MEDLINE
L20   0 FILE HCAPLUS
L21   0 FILE BIOSIS
L22   0 FILE EMBASE
L23   0 FILE JICST-EPLUS
L24   0 FILE WPIDS

TOTAL FOR ALL FILES
L25   0 L18 NOT L10

=> (l1 or cub domain or extracellular domain or ecd) and (novel human protein? or
nhps)
(L1 IS NOT A RECOGNIZED COMMAND
The previous command name entered was not recognized by the system.
For a list of commands available to you in the current file, enter
"HELP COMMANDS" at an arrow prompt (=>).

=> s (l1 or cub domain or extracellular domain or ecd) and (novel human protein? or
nhps)
L26   1 FILE MEDLINE
L27   7 FILE HCAPLUS
L28   1 FILE BIOSIS
L29   1 FILE EMBASE
L30   0 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'
L31   1 FILE WPIDS

TOTAL FOR ALL FILES
L32   11 (L1 OR CUB DOMAIN OR EXTRACELLULAR DOMAIN OR ECD) AND (NOVEL
             HUMAN PROTEIN? OR NHPS)

=> s l32 not (l10 or l18)
L33   1 FILE MEDLINE
L34   6 FILE HCAPLUS
L35   1 FILE BIOSIS
L36   1 FILE EMBASE
L37   0 FILE JICST-EPLUS
L38   0 FILE WPIDS

Searched by: Mary Hale 308-4258 CM-1 1E01

TOTAL FOR ALL FILES
L39   9 L32 NOT (L10 OR L18)

=> dup rem 139
PROCESSING COMPLETED FOR L39
L40   6 DUP REM L39 (3 DUPLICATES REMOVED)

=> d cbib abs 1-6

L40 ANSWER 1 OF 6 HCAPLUS COPYRIGHT 2002 ACS
2002:107557   Document No. 136:162371 Cloning and characterization of
     novel human protein kinase family members
     32374 and 18431 and their therapeutic uses. Meyers, Rachel;
     Kapeller-Libermann, Rosana; Silos-Santiago, Immaculada (Millennium
     Pharmaceuticals, Inc., USA). PCT Int. Appl. WO 2002010401 A2 20020207,
     141 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG,
     BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI,
     GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
     LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT,
     RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN,
     YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF,
     CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML,
     MR, NE, NL, PT, SE, SN, TD, TG, TR. (English). CODEN: PIXXD2.
     APPLICATION: WO 2001-US23653 20010727. PRIORITY: US 2000-PV221543
     20000728.
AB   The invention provides isolated nucleic acids mols., designated 32374 or
     18431 nucleic acid mols., which encode novel protein kinase family
     members. The invention also provides antisense nucleic acid mols.,
     recombinant expression vectors contg. 32374 or 18431 nucleic acid mols.,
     host cells into which the expression vectors have been introduced, and
     nonhuman transgenic animals in which a 32374 or 18431 gene has been
     introduced or disrupted. Their putative function domains are analyzed and
     their gene expression profiles are provided. The invention still further
     provides isolated 32374 or 18431 proteins, fusion proteins, antigenic
     peptides and anti-32374 or -18431 antibodies. Diagnostic methods
     utilizing compns. of the invention are also provided.

L40 ANSWER 2 OF 6 HCAPLUS COPYRIGHT 2002 ACS
2002:10647   Document No. 136:80918 Protein and cDNA sequences of a novel
     human **CUB domain**-containing protein sequence homolog
     and uses thereof. Kapeller-Libermann, Rosana (Millennium Pharmaceuticals,
     Inc., USA). PCT Int. Appl. WO 2002000843 A2 20020103, 107 pp. DESIGNATED
     STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH,
     CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM,
     HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,
     LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG,
     SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ,
     BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY,
     DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE,
     SN, TD, TG, TR. (English). CODEN: PIXXD2. APPLICATION: WO 2001-US20055
     20010621. PRIORITY: US 2000-PV213963 20000623.
AB   The invention provides protein and cDNA sequences of a **novel
     human protein,** designated 56739, which has sequence
     homol. with CUB family members. The invention also provides antisense
     nucleic acid mols., recombinant expression vectors contg. 56739 nucleic
     acid mols., host cells into which the expression vectors have been
     introduced, and nonhuman transgenic animals in which a 56739 gene has been
     introduced or disrupted. The invention still further provides isolated
     56739 proteins, fusion proteins, antigenic peptides and anti-56739
     antibodies. Diagnostic methods utilizing compns. of the invention are
     also provided.

2001:526099   Document No. 135:117953   cDNA and protein sequence of
     novel human protein TANGO416 and TANGO457 and
     their uses in drug screening, therapeutics and diagnosis. Fraser,
     Christopher C.; Kirst, Susan J. (Millennium Pharmaceuticals, Inc., USA).
     PCT Int. Appl. WO 2001051514 A1 20010719, 217 pp. DESIGNATED STATES: W:
     AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU,
     CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN,
     IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK,
     MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
     TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ,
     TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA,
     GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG, TR.
     (English). CODEN: PIXXD2. APPLICATION: WO 2001-US485 20010105.
     PRIORITY: US 2000-479249 20000107; US 2000-559497 20000427.

AB   The invention provides cDNA and protein sequence of two **novel
     human protein** TANGO416 and TANGO457. The protein
     TANGO416 was cloned from human fetal spleen and has high homol. with
     murine protocadherin. Both proteins are transmembrane proteins and contg.
     extracellular, transmembrane and intracellular domain. These nucleic and
     proteins are useful for diagnosis, prevention, and therapy of a no. of
     human and other animal disorders. The invention also provides antisense
     nucleic acid mols., expression vectors contg. the nucleic acid mols. of
     the invention, host cells into which the expression vectors have been
     introduced, and non-human transgenic animals in which a nucleic acid mol.
     of the invention has been introduced or disrupted. The invention still
     further provides isolated polypeptides, fusion polypeptides, antigenic
     peptides and antibodies. Diagnostic, screening, and therapeutic methods
     using compns. of the invention are also provided. The nucleic acids and
     polypeptides of the present invention are useful as modulating agents in
     regulating a variety of cellular processes.

2001:300759   Document No. 134:321609   Protein and cDNA sequences of human
     secretory proteins and uses thereof in diagnosis, therapy and drug
     screening. Turner, C. Alexander, Jr.; Donoho, Gregory; Nehls, Michael;
     Hilbun, Erin; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics
     Incorporated, USA). PCT Int. Appl. WO 2001029084 A2 20010426, 63 pp.
     DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ,
     CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM,
     HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,
     LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG,
     SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY,
     KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE,
     DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN,
     TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US28803
     20001018. PRIORITY: US 1999-PV160106 19991018; US 1999-PV162547 19991029.

AB   This invention provides protein and cDNA sequences for newly identified
     human proteins, designated **NHPs**, which shares structure motifs
     typical of human secreted proteins that share structural similarity with
     animal chordins, NEL protein, and thrombospondin. Also disclosed is
     another **novel human protein** that shares
     structural similarity with animal protein that contains **CUB
     domain**. In one embodiment, the invention relates to diagnostic
     assays for detecting diseases assocd. with inappropriate NHP activity or
     levels. Also disclosed are methods for utilizing NHP in drug screening
     assays and in therapy directed against diseases assocd. with inappropriate
     NHP activity or levels.

1998:608707   Document No. 129:215729   Cloning of cDNA for **novel**

human **proteins** belong to tumor necrosis factor receptor
superfamily. Tada, Hideaki; Konishi, Mikio; Fukushima, Daikichi (Ono
Pharmaceutical Co., Ltd., Japan). PCT Int. Appl. WO 9838304 A1 19980903,
54 pp. DESIGNATED STATES: W: JP, KR, US; RW: AT, BE, CH, DE, DK, ES, FI,
FR, GB, GR, IE, IT, LU, MC, NL, PT, SE. (Japanese). CODEN: PIXXD2.
APPLICATION: WO 1998-JP799 19980226. PRIORITY: JP 1997-43143 19970227.

AB   The cDNA for proteins OAF065.alpha. and OAF065.beta. are isolated from
human stroma cell line HAS303. OAF065.alpha. consists of 416 and
OAF065.beta. of 423 amino acids, with identical N-terminal 1-415 residues.
The C-terminal Glu-Ala of OAF065.alpha. is substituted with 8 amino acid
residues in OAF065.beta.. Both protein contains a Cys-rich domain,
commonly seen in the **extracellular domain** of the TNF
receptor superfamily. Claimed are methods of recombinant prepn. of the
proteins, (monoclonal) antibody to the proteins, oligonucleotide
hybridizing to the cDNA, and therapeutics contg. the protein and/or the
antibodies.

L40 ANSWER 6 OF 6   MEDLINE   DUPLICATE 1
96394649 Document Number: 96394649.   PubMed ID: 8798755.   Cloning and
characterization of islet cell antigen-related protein-tyrosine
phosphatase (PTP), a novel receptor-like PTP and autoantigen in
insulin-dependent diabetes. Cui L; Yu W P; DeAizpurua H J; Schmidli R S;
Pallen C J. (Cell Regulation Laboratory, Institute of Molecular and Cell
Biology, National University of Singapore, 10 Kent Ridge Crescent,
Singapore 0511, Republic of Singapore.) JOURNAL OF BIOLOGICAL CHEMISTRY,
(1996 Oct 4) 271 (40) 24817-23. Journal code: 2985121R. ISSN: 0021-9258.
Pub. country: United States. Language: English.

AB   Cloning of the cDNA encoding a **novel human**
protein- tyrosine phosphatase (PTP) called islet cell
antigen-related PTP (IAR) predicts a receptor-like molecule with an
extracellular domain of 614 amino acids containing a
hydrophobic signal peptide, one potential N-glycosylation site, and an
RGDS peptide which is a possible adhesive recognition sequence. The
376-amino acid intracellular region contains a single catalytic domain.
Recombinant IAR polypeptide has phosphatase activity. Northern blot
analysis shows tissue-specific expression of two IAR transcripts of 5.5
and 3. 7 kilobases, which are most abundant in brain and pancreas. The IAR
PTP is homologous in its intracellular region to IA-2, a putative PTP that
is an insulin-dependent diabetes mellitus (IDDM) autoantigen. IAR is also
reactive with IDDM patient sera. IAR and IA-2 may distinguish different
populations of IDDM autoantibodies since they identify overlapping but
nonidentical sets of IDDM patients. Thus IAR is likely to be an islet cell
antigen useful in the preclinical screening of individuals for risk of
IDDM.

=> s (l1 or cub domain or extracellular domain or ecd) and (agonist or antagonist)
L41   611 FILE MEDLINE
L42   485 FILE HCAPLUS
L43   408 FILE BIOSIS
L44   359 FILE EMBASE
L45   36 FILE JICST-EPLUS

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'CUB-DOMAIN-CONTAINING PROTEIN
RP43 (RIFTIA PACHYPTILA/CN'
L46   203 FILE WPIDS

TOTAL FOR ALL FILES
L47   2102 (L1 OR CUB DOMAIN OR EXTRACELLULAR DOMAIN OR ECD) AND (AGONIST

Searched by: Mary Hale 308-4258 CM-1 1E01

OR ANTAGONIST)

```
=> s l47 and (l3 or discoidin(w)(receptor! or domain))
L48          0 FILE MEDLINE
L49          1 FILE HCAPLUS
L50          0 FILE BIOSIS
L51          0 FILE EMBASE
L52          0 FILE JICST-EPLUS
```

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'   -

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'

EXCEEDS MAXIMUM FIELD LENGTH, WILL BE SEARCHED AS 'DISCOIDIN DOMAIN RECEPTOR KINASE
(MUS MUSCULUS STRAIN/CN'
L53   0 FILE WPIDS
```

TOTAL FOR ALL FILES
L54   1 L47 AND (L3 OR DISCOIDIN(W)(RECEPTOR! OR DOMAIN))

=> d cbib abs

L54   ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2002 ACS
1998:71221   Document No. 128:136859 Soluble ErbB receptor
     extracellular domain fusion proteins their uses in
     antagonization of growth factors. Fizpatrick, Vincent Danial; Sliwkowski,
     Mark; Vandlen, Richard L. (Genentech, Inc., USA). PCT Int. Appl. WO
     9802540 A1 19980122, 58 pp. DESIGNATED STATES: W: AL, AM, AT, AU, AZ,
     BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH,
     HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG,
     MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
     TR, TT, UA, UG, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW:
     AT, BE, BF, BJ, CF, CG, CH, CI, CM, DE, DK, ES, FI, FR, GA, GB, GR, IE,
     IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN:
     PIXXD2. APPLICATION: WO 1997-US11825 19970708. PRIORITY: US 1996-21640
     19960712; US 1997-798326 19970210.
AB   Sol. heteromultimeric immunoadhesin fusion proteins that contain the
     extracellular ligand-binding domains of ErbB receptors oligomerized via
     the multimerization domains of Igs are described for use as
     antagonists of neuregulins in the control of cell proliferation,
     e.g. in cancers. Specifically, the **extracellular**
     domains of ErbB2 and ErbB3 or ErbB2 and ErbB4 are used in
     combination. The prepn. ErbB/IgG fusion proteins and the characterization
     of their heregulin binding properties is reported.


```
=> s donoho, g?/au,in;s turner, c?/au;s nehls, m?/au,in;s friedrich, g?/au,in
'IN' IS NOT A VALID FIELD CODE
L55          6 FILE MEDLINE
L56      .  44 FILE HCAPLUS
L57          9 FILE BIOSIS
'IN' IS NOT A VALID FIELD CODE
L58          5 FILE EMBASE
L59          0 FILE JICST-EPLUS
```

Searched by: Mary Hale 308-4258 CM-1 1E01

```
L60          36 FILE WPIDS

TOTAL FOR ALL FILES
L61         100 DONOHO, G?/AU,IN


L62        1091 FILE MEDLINE
L63        1428 FILE HCAPLUS
L64        1198 FILE BIOSIS
L65         772 FILE EMBASE
L66           7 FILE JICST-EPLUS
L67         212 FILE WPIDS

TOTAL FOR ALL FILES
L68        4708 TURNER, C?/AU


'IN' IS NOT A VALID FIELD CODE
L69          38 FILE MEDLINE
L70          77 FILE HCAPLUS
L71          42 FILE BIOSIS
'IN' IS NOT A VALID FIELD CODE
L72          30 FILE EMBASE
L73           0 FILE JICST-EPLUS
L74          39 FILE WPIDS

TOTAL FOR ALL FILES
L75         226 NEHLS, M?/AU,IN


'IN' IS NOT A VALID FIELD CODE
L76         158 FILE MEDLINE
L77         337 FILE HCAPLUS
L78         163 FILE BIOSIS
'IN' IS NOT A VALID FIELD CODE
L79         167 FILE EMBASE
L80           1 FILE JICST-EPLUS
L81         105 FILE WPIDS

TOTAL FOR ALL FILES
L82         931 FRIEDRICH, G?/AU,IN

=> s 161 and 168 and 175 and 182
L83           0 FILE MEDLINE
L84          10 FILE HCAPLUS
L85           0 FILE BIOSIS
L86           0 FILE EMBASE
L87           0 FILE JICST-EPLUS
L88          10 FILE WPIDS

TOTAL FOR ALL FILES
L89          20 L61 AND L68 AND L75 AND L82

=> dup rem 189
PROCESSING COMPLETED FOR L89
L90          10 DUP REM L89 (10 DUPLICATES REMOVED)

=> d 1-10 cbib abs

L90  ANSWER 1 OF 10  HCAPLUS  COPYRIGHT 2002 ACS       DUPLICATE 1
2001:360153   Document No. 134:363350  Characterization and applications of
     novel human proteases. Donoho, Gregory; Hilbun, Erin; Turner,
```

C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001034779 A2 20010517, 34 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG, TR. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US31179 20001109. PRIORITY: US 1999-PV165260 19991112.

AB The cDNA sequences and corresponding deduced amino acid sequences of novel human proteases (NHPs) are disclosed. The NHP genes were obtained from a human testis cDNA library using probes and/or primers generated from human gene trapped sequence tags. The described NHPs share significant similarity to a variety of mammalian proteases, and esp. carboxypeptidase A. The NHPs can be used in therapeutic, diagnostic, and pharmacogenomic applications.

L90 ANSWER 2 OF 10 HCAPLUS COPYRIGHT 2002 ACS   DUPLICATE 2
2001:360152   Document No. 134:363349 Characterization of novel human ATPase homologs. **Donoho, Gregory**; Turner, C. Alexander, Jr.; Hilbun, Erin; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001034778 A2 20010517, 45 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG, TR. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US31110 20001109. PRIORITY: US 1999-PV164624 19991110.

AB The present invention relates to the discovery, identification, and characterization of novel human proteins which share structural similarity with calcium transporting ATPases and aminophospholipid transporters. The cDNA sequences and the corresponding deduced amino acid sequences of the human ATPase homologs are disclosed. The human ATPase homologs can be used in therapeutic, diagnostic, and pharmacogenomic applications.

L90 ANSWER 3 OF 10 HCAPLUS COPYRIGHT 2002 ACS   DUPLICATE 3
2001:338573   Document No. 134:348997 Protein and cDNA sequences of human transport proteins and uses thereof in diagnosis, therapy and drug screening. Harras, Marie; **Donoho, Gregory; Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn;** Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001032706 A2 20010510, 59 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US29852 20001031. PRIORITY: US 1999-PV163018 19991102.

AB This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which share structural similarity with mammalian multi-drug resistance (MDR) proteins and cellular transporters. In one embodiment, the invention relates to diagnostic assays for detecting diseases associd. with inappropriate NHP activity or levels.

Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels. Seq ID No. 47 and 48 were claimed, but information was not provided.

L90   ANSWER 4 OF 10   HCAPLUS   COPYRIGHT 2002 ACS   DUPLICATE 4
    2001:300872   Document No. 134:321614   Protein and cDNA sequences of a novel human CUB-domain containing protein and uses thereof in diagnosis, therapy and drug screening. **Donoho, Gregory**; Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001029219 A1 20010426, 33 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US28798 20001008. PRIORITY: US 1999-PV160285 19991019; US 2000-PV183583 20000218.
AB   This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which shares structural similarity with animal CUB domain proteins, coagulation factors V and XIII, milk fat globule-EGF factor 8, 5 transcriptional repressor AE-binding protein-1, and neuropilins 1 and 2 (which, like the presently described protein, contain both CUB and discoidin domains). While NHP shares sequence homol. with other CUB-domain contg. proteins, its primary sequence is unique. Its expression is detected in human cell lines, and human prostate, pituitary, fetal brain, brain, thymus, spleen, lymph node, trachea, kidney, fetal liver, thyroid, adrenal gland, salivary gland, stomach, small intestine, colon, muscle, heart, mammary gland, adipose, skin, esophagus, bladder, cervix, rectum, and testis cells. In one embodiment, the invention relates to diagnostic assays for detecting diseases assocd. with inappropriate NHP activity or levels. Also disclosed are methods for utilizing NHP in drug screening assays and in therapy directed against diseases assocd. with inappropriate NHP activity or levels.

L90   ANSWER 5 OF 10   HCAPLUS   COPYRIGHT 2002 ACS   DUPLICATE 5
    2001:247518   Document No. 134:276537   Protein and cDNA sequences of human carboxypeptidases and uses thereof in diagnosis, therapy and drug screening. **Donoho, Gregory**; Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001023588 A1 20010405, 36 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26876 20000929. PRIORITY: US 1999-PV156685 19990929.
AB   This invention provides protein and cDNA sequences for newly identified human proteins, designated NHPs, which shares substantial sequence homol. with animal proteases, esp. carboxypeptidases. While NHP shares sequence homol. with other carboxypeptidases, its primary sequence is unique. Its expression is detected in various human tissues including brain, pituitary, spinal cord, thymus, spleen, lymph node, bone marrow, trachea, lung, kidney, prostate, testis, thyroid, adrenal gland, stomach, small intestine colon, skeletal muscle, uterus, mammary gland, bladder, cervix cells, and gene trapped human cells. In one embodiment, the invention

Searched by: Mary Hale 308-4258 CM-1 1E01

relates to diagnostic assays for detecting diseases assocd. with
inappropriate NHP activity or levels. Also disclosed are methods for
utilizing NHP in drug screening assays and in therapy directed against
diseases assocd. with inappropriate NHP activity or levels.

L90   ANSWER 6 OF 10   HCAPLUS   COPYRIGHT 2002 ACS   DUPLICATE 6
2001:247510   Document No. 134:261891 Protein and cDNA sequences of human
    serine/threonine protein kinase and uses thereof in diagnosis, therapy and
    drug screening. **Donoho, Gregory**; Turner, C. Alexander, Jr.;
    Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T.
    (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001023579 A1
    20010405, 38 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA,
    BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI,
    GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
    LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT,
    RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU,
    ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG,
    CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR,
    NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO
    2000-US26621 20000927. PRIORITY: US 1999-PV156511 19990928.
AB   This invention provides protein and cDNA sequences for newly identified
    human proteins, designated NHPs, which shares substantial sequence homol.
    with animal kinases, and more particular serine/threonine protein kinases.
    While NHP shares sequence homol. with other serine/threonine protein
    kinases, its primary sequence is unique. Its expression is detected in
    various human tissues including brain, pituitary, spinal cord, spleen,
    trachea, kidney, prostate, testis, adrenal gland cells, and gene trapped
    human cells. In one embodiment, the invention relates to diagnostic
    assays for detecting diseases assocd. with inappropriate NHP activity or
    levels. Also disclosed are methods for utilizing NHP in drug screening
    assays and in therapy directed against diseases assocd. with inappropriate
    NHP activity or levels.

L90   ANSWER 7 OF 10   HCAPLUS   COPYRIGHT 2002 ACS   DUPLICATE 7
2001:247499   Document No. 134:261870 Protein and cDNA sequences of a novel
    human organic anion transport protein. **Turner, C. Alexander, Jr.**
    ; **Donoho, Gregory**; Wattler, Frank; **Nehls, Michael;**
    Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon
    Genetics Incorporated, USA). PCT Int. Appl. WO 2001023566 A1 20010405, 31
    pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR,
    BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE,
    GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,
    LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD,
    SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM,
    AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM,
    CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT,
    SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26630
    20000927. PRIORITY: US 1999-PV156161 19990927.
AB   This invention provides protein and cDNA sequences for newly identified
    human proteins, designated NHPs, which shares substantial sequence homol.
    with animal org. anion, and more particularly prostaglandin, transporter
    proteins. While NHP shares sequence homol. with other org. anion
    transport proteins, its primary sequence is unique. Its expression is
    detected in various human tissues including brain, pituitary, spinal cord,
    lymph node, trachea, heart, adipose, skin, pericardium, hypothalamus
    cells, and gene trapped human cells. In one embodiment, the invention
    relates to diagnostic assays for detecting diseases assocd. with
    inappropriate NHP activity or levels. Also disclosed are methods for
    utilizing NHP in drug screening assays and in therapy directed against
    diseases assocd. with inappropriate NHP activity or levels.

L90   ANSWER 8 OF 10   HCAPLUS   COPYRIGHT 2002 ACS   DUPLICATE 8

2001:229020   Document No. 134:262848 Cloning and characterization of human endothelin converting enzyme-like proteins. **Donoho, Gregory;** Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001021773 A2 20010329, 31 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26047 20000922. PRIORITY: US 1999-PV156102 19990924; US 2000-PV176689 20000118.

AB   Novel human endothelin converting enzyme-like proteins of 255 and 883 amino acids in length and cDNAs encoding these proteins are disclosed. The endothelin converting enzyme-like proteins and encoding nucleic acids can be used in therapeutic, diagnostic, and pharmacogenomic applications.

2001:228919   Document No. 134:247996 Protein and cDNA sequences for a novel human protease inhibitor-like protein NHP and use thereof. **Donoho, Gregory;** Turner, C. Alexander, Jr.; Wattler, Frank; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001021651 A2 20010329, 29 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US26048 20000922. PRIORITY: US 1999-PV156101 19990924.

AB   The invention provides protein and cDNA sequences for a novel human protease inhibitor-like protein NHP that can be used in therapeutic, diagnostic, and pharmacogenomic applications.

2001:168158   Document No. 134:189017 Protein and cDNA sequences of human calcium dependent protease homologs, and uses thereof in therapy, diagnosis and drug screening. **Donoho, Gregory;** Turner, C. Alexander, Jr.; Nehls, Michael; Friedrich, Glenn; Zambrowicz, Brian; Sands, Arthur T. (Lexicon Genetics Incorporated, USA). PCT Int. Appl. WO 2001016336 A1 20010308, 39 pp. DESIGNATED STATES: W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM; RW: AT, BE, BF, BJ, CF, CG, CH, CI, CM, CY, DE, DK, ES, FI, FR, GA, GB, GR, IE, IT, LU, MC, ML, MR, NE, NL, PT, SE, SN, TD, TG. (English). CODEN: PIXXD2. APPLICATION: WO 2000-US24062 20000901. PRIORITY: US 1999-PV152057 19990902.

AB   The invention provides protein and cDNA sequences of a novel class of human calcium dependent protease homologs which share structural similarity with animal calcium dependent proteases, or calpains. The novel calcium dependent protease homolog encoding cDNAs were obtained from human gene trap clones and human cDNA libraries. The invention further relates to the uses of calcium dependent protease homolog in therapy, diagnosis and drug screening.

```
=> log y
```

COST IN U.S. DOLLARS	SINCE FILE ENTRY	TOTAL SESSION
FULL ESTIMATED COST	71.52	120.70

DISCOUNT AMOUNTS (FOR QUALIFYING ACCOUNTS)	SINCE FILE ENTRY	TOTAL SESSION
CA SUBSCRIBER PRICE	-10.53	-10.53

STN INTERNATIONAL LOGOFF AT 16:04:39 ON 09 JUL 2002

Searched by: Mary Hale 308-4258 CM-1 1E01